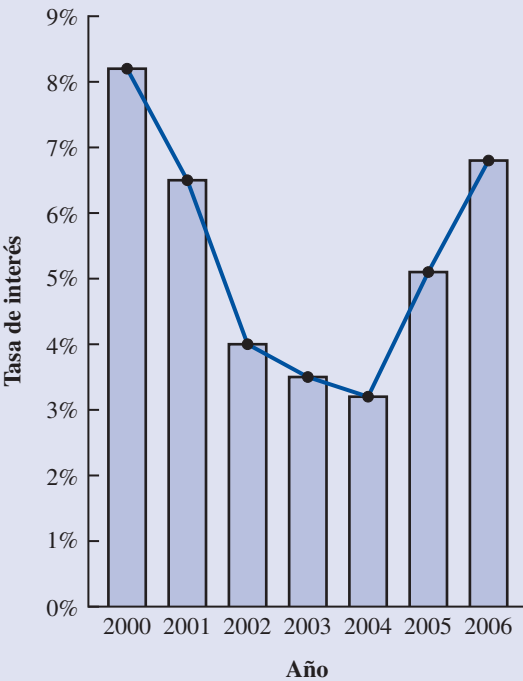
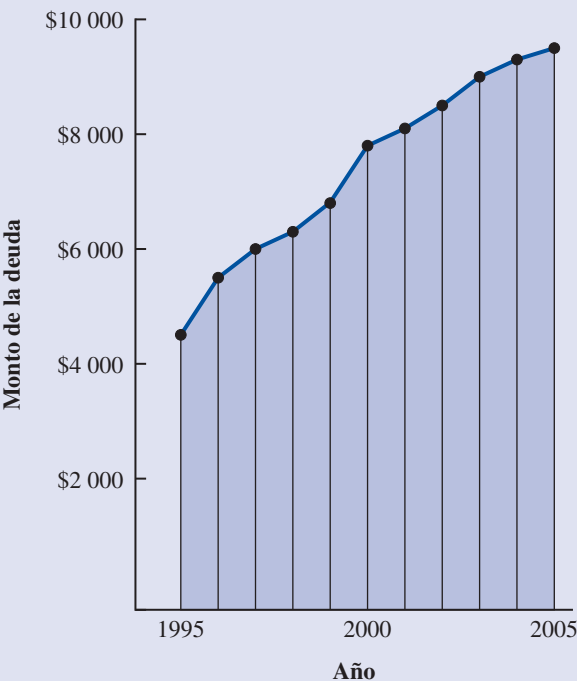


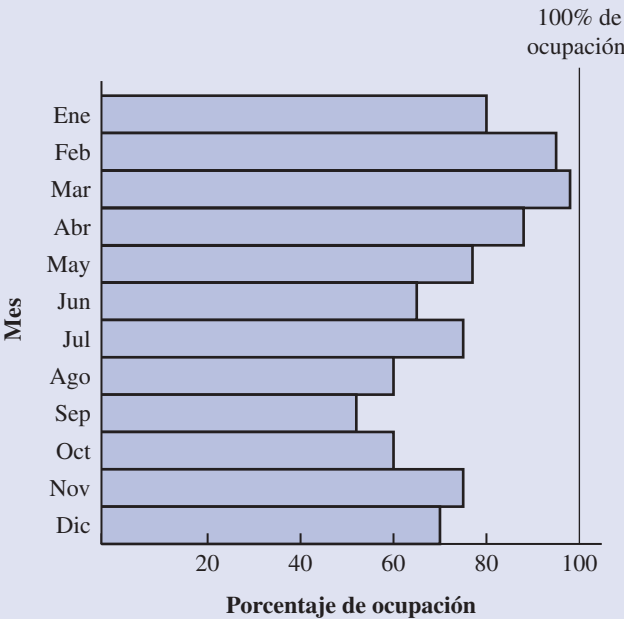
FIGURA 1.2 DIVERSAS GRÁFICAS DE DATOS DE SERIES DE TIEMPO



(A) Tasas de interés en los Stafford Loans para estudiantes



(B) Adeudo promedio en tarjetas de crédito por hogar



(C) Tasas de ocupación en hoteles de Florida del sur

## LA ESTADÍSTICA en LA PRÁCTICA

### LA EMPRESA COLGATE-PALMOLIVE\* NUEVA YORK, NUEVA YORK

La empresa Colgate-Palmolive empezó en la Ciudad de Nueva York en 1806 como una pequeña tienda de jabones y velas. Hoy, Colgate-Palmolive emplea más de 4000 personas que trabajan en 200 países y territorios del mundo. Aunque es más conocida por sus marcas Colgate, Palmolive, Ajax y Fab, la empresa comercializa los productos Mennen, Hill's Science Diet y Hill's Prescription Diet.

La empresa Colgate-Palmolive aplica la estadística en su programa de aseguramiento de la calidad en los detergentes caseros para la ropa. Le interesa la satisfacción del cliente con la cantidad de detergente en los paquetes. Todos los paquetes de cierto tamaño se llenan con la misma cantidad de detergente en peso, aunque el volumen del detergente varía de acuerdo con la densidad del polvo detergente. Por ejemplo, si la densidad del detergente es alta, se necesita una cantidad menor de detergente para tener el peso señalado en el paquete. El resultado es que cuando el cliente abre el paquete le parece que no ha sido bien llenado.

Para controlar el problema del peso del polvo de detergente, se han establecido límites en el nivel aceptable de la densidad del polvo. Con periodicidad se toman muestras estadísticas y se mide la densidad de la muestra de polvo. Los resúmenes de los datos se les proporcionan a los operarios para que de ser necesario lleven a cabo acciones correctivas, de manera que la densidad se mantenga dentro de las especificaciones de calidad establecidas.

En la tabla y figura adjuntas se presentan una distribución de frecuencia y un histograma obtenidos con 150 muestras tomadas en una semana. Densidades mayores a 0.40 son inaceptablemente altas. De acuerdo con la distribución de frecuencia y al histograma la operación satisface los lineamientos de calidad ya que todas las densidades son menores o iguales a 0.40. A la vista de estos resúmenes estadísticos los directivos estarán satisfechos con la calidad del proceso de producción de detergente.

En este capítulo se estudiarán métodos tabulares y gráficos de la estadística descriptiva como distribuciones de frecuencia, gráficas de barras, histogramas, diagramas de tallo y hoja, tabulaciones cruzadas y otros. El objeto de



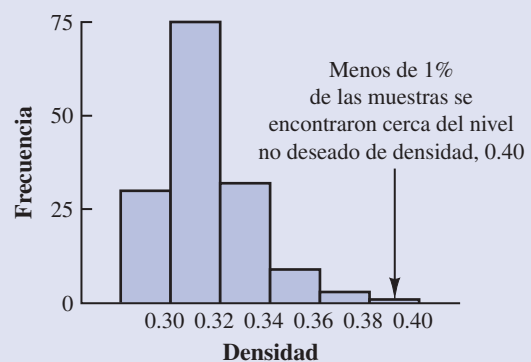
Los resúmenes estadísticos ayudan a mantener la calidad de estos productos de Colgate-Palmolive  
© Joe Higgins/South Western.

estos métodos es resumir los datos de manera que sean entendibles e interpretables con facilidad.

#### Distribución de frecuencia de los datos de densidad

Densidad	Frecuencia
0.29–0.30	30
0.31–0.32	75
0.33–0.34	32
0.35–0.36	9
0.37–0.38	3
0.39–0.40	1
Total	150

#### Histograma de los datos de densidad



\*Los autores agradecen a William R. Fawle, director de aseguramiento de la calidad de la empresa Colgate-Palmolive por proporcionarles este artículo para *La estadística en la práctica*.

TABLA 2.2

DISTRIBUCIÓN DE  
FRECUENCIA DE  
LAS VENTAS DE  
REFRESCO

Refresco	Frecuencia
Coke Classic	19
Diet Coke	8
Dr. Pepper	5
Pepsi	13
Sprite	5
Total	50

Para elaborar una distribución de frecuencia con estos datos, se cuenta el número de veces que aparece cada refresco en la tabla 2.1. La Coca cola clásica (Coke Classic) aparece 19 veces, la Coca cola de dieta (Diet Coke) 8 veces, Dr. Pepper 5 veces, Pepsi 13 veces y Sprite 5 veces. Esto queda resumido en la distribución de frecuencia de la tabla 2.2.

Esta distribución de frecuencia proporciona un resumen de cómo se distribuyeron las 50 ventas entre los cinco refrescos. El resumen aporta más claridad que los datos originales de la tabla 2.1. Al observar esta distribución de frecuencia, es claro que Coca cola clásica es el refresco que más se vende, Pepsi el segundo, Coca cola de dieta el tercero y Sprite y Dr. Pepper están empatados en el cuarto lugar. La distribución de frecuencia resume la información sobre la popularidad de los cinco refrescos.

## Distribuciones de frecuencia relativa y de frecuencia porcentual

En una distribución de frecuencia se aprecia el número (frecuencia) de los elementos de cada una de las diversas clases disjuntas. Sin embargo, con frecuencia lo que interesa es la proporción o porcentaje de elementos en cada clase. La *frecuencia relativa* de una clase es igual a la parte o proporción de los elementos que pertenecen a cada clase. En un conjunto de datos, en el que hay  $n$  observaciones, la frecuencia relativa de cada clase se determina como sigue:

### FRECUENCIA RELATIVA

$$\text{Frecuencia relativa de una clase} = \frac{\text{Frecuencia de la clase}}{n} \quad (2.1)$$

La *frecuencia porcentual* de una clase es la frecuencia relativa multiplicada por 100.

Una **distribución de frecuencia relativa** da un resumen tabular de datos en el que se muestra la frecuencia relativa de cada clase. Una **distribución de frecuencia porcentual** da la frecuencia porcentual de los datos de cada clase. En la tabla 2.3 se presenta una distribución de frecuencia relativa y una distribución de frecuencia porcentual de los datos de los refrescos. En esta tabla se observa que la frecuencia relativa de la Coca cola clásica es  $19/50 = 0.38$ , la de la Coca cola de dieta es  $8/50 = 0.16$ , etc. En la distribución de frecuencia porcentual, se muestra que 38% de las ventas fueron de Coca cola clásica, 16% de Coca cola de dieta, etc. También resulta que  $38\% + 26\% + 16\% = 80\%$  de las ventas fueron de los tres refrescos que más se venden.

## Gráficas de barra y gráficas de pastel

Una **gráfica de barras** o un diagrama de barras, es una gráfica para representar los datos cualitativos de una distribución de frecuencia, de frecuencia relativa o de frecuencia porcentual. En uno de los ejes de la gráfica (por lo general en el horizontal), se especifican las etiquetas empleadas para las clases (categorías). Para el otro eje de la gráfica (el vertical) se usa una escala para

TABLA 2.3 DISTRIBUCIONES DE FRECUENCIA RELATIVA Y FRECUENCIA PORCENTUAL DE LAS VENTAS DE REFRESCOS

Refresco	Frecuencia relativa	Frecuencia porcentual
Coke Classic	0.38	38
Diet Coke	0.16	16
Dr. Pepper	0.10	10
Pepsi	0.26	26
Sprite	0.10	10
Total	1.00	100



## 2-3 Histogramas

**Concepto clave** En la sección 2-2 se presentó la distribución de frecuencias como una herramienta para resumir y conocer la naturaleza de la distribución de un conjunto de datos grande. Esta sección presenta el *histograma* como una gráfica muy importante para describir la naturaleza de la distribución. Puesto que muchos programas de cómputo estadísticos y calculadoras generan histogramas de forma automática, no es tan importante dominar los procedimientos mecánicos para construirlos. En cambio, debemos enfocarnos en *comprender* lo que ganaríamos al examinar histogramas. En particular, debemos desarrollar la habilidad de observar un histograma y entender la naturaleza de la distribución de los datos.

### Definición

Un **histograma** es una gráfica de barras donde la escala horizontal representa clases de valores de datos y la escala vertical representa frecuencias. Las alturas de las barras corresponden a los valores de frecuencia; en tanto que las barras se dibujan de manera adyacente (sin huecos entre sí).

El primer paso para la construcción de un histograma es la creación de una tabla de distribución de frecuencias. El histograma es, básicamente, una versión gráfica de dicha tabla. Observe la figura 2-2, que incluye el histograma correspondiente a la distribución de frecuencias de la tabla 2-2 de la sección anterior.

En la escala horizontal se marca cada barra del histograma con su frontera de clase inferior a la izquierda, y su frontera de clase superior a la derecha, como se observa en la figura 2-2. En vez de utilizar las fronteras de clase a lo largo del eje horizontal, a menudo es más práctico el uso de los valores de la marca de clase en

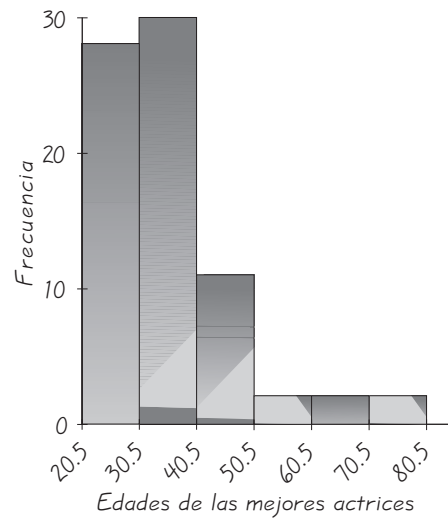
**Tabla del ejercicio 27**

Número de valores	Número ideal de clases
16–22	5
23–45	6
	7
	8
	9
	10
	11
	12

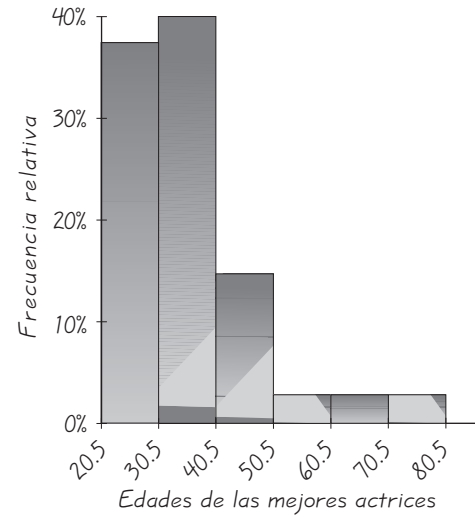


### Datos faltantes

Es común que las muestras carezcan de algunos datos. Los datos faltantes caen en dos categorías generales: **1.** valores faltantes que resultan de causas aleatorias no relacionadas con los valores de los datos, y **2.** valores faltantes que resultan de causas que no son aleatorias. Algunas de las causas aleatorias son factores como la anotación incorrecta de valores muestrales o la pérdida de resultados de encuesta. Este tipo de valores faltantes a menudo puede despreciarse, ya que no ocultan de manera sistemática algunas características que podrían afectar los resultados significativamente. Es difícil enfrentarse a valores faltantes que no se deben al azar. Por ejemplo, los resultados del análisis del ingreso podrían verse seriamente afectados, si la gente con ingresos muy altos se rehúsa a proporcionar los valores por temor a las auditorías fiscales. Tales ingresos muy altos faltantes no deben descartarse; más bien, se debe realizar otra investigación para identificarlos.



**Figura 2-2 Histograma**



**Figura 2-3 Histograma de frecuencias relativas**

el centro de las barras correspondientes. El uso de los valores de la marca de clase es muy común en los programas de cómputo que generan histogramas de manera automática.

**Escala horizontal:** Utilice las fronteras de clase o las marcas de clase.

**Escala vertical:** Utilice las frecuencias de clase.

Antes de construir un histograma a partir de una distribución de frecuencias terminada, debemos hablar de las escalas que se emplean en los ejes vertical y horizontal. La frecuencia máxima (o el siguiente número más alto conveniente) debería sugerir un valor para la parte superior de la escala vertical; 0 tiene que estar en la parte inferior. En la figura 2-2 diseñamos la escala vertical para que corriera de 0 a 30. La escala horizontal debe subdividirse de tal forma que permita que todas las clases se ajusten bien. De manera ideal, es necesario que intentemos seguir la regla general de que la altura vertical del histograma debe medir alrededor de tres cuartas partes de la anchura total. A ambos ejes se les asignan etiquetas claras.

### Histograma de frecuencias relativas

Un **histograma de frecuencias relativas** tiene la misma forma y escala horizontal que un histograma, pero la escala vertical está marcada con las frecuencias relativas en vez de las frecuencias reales, como se observa en la figura 2-3.

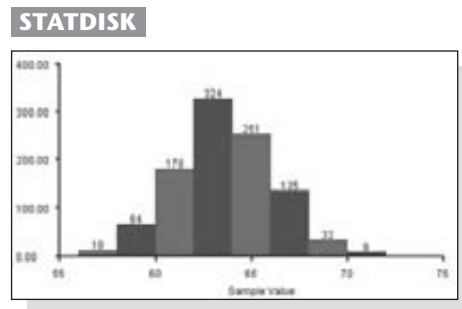
### Pensamiento crítico: interpretación de histogramas

Recuerde que el objetivo no es la simple construcción de un histograma, sino *entender* algo acerca de los datos. Analice el histograma para ver qué es posible aprender acerca de “CVDVT”: el centro de los datos, la variación (que se estudiará con detalle en la sección 3-3), la forma de la distribución y la existencia o ausencia de valores extremos (valores que se encuentran lejos de los demás). Al examinar

la figura 2-2, vemos que el histograma se centra alrededor del 35, que los valores varían aproximadamente desde 21 hasta 80, y que la forma de la distribución está más cargada hacia la izquierda, lo cual significa que las actrices que ganan un Óscar tienden a ser desproporcionadamente más jóvenes, y que a pocas actrices mayores se les otorga ese premio.

**Distribución normal** En la sección 2-2 señalamos que el uso del término “normal” tiene un significado especial en la estadística, el cual difiere del significado que generalmente se le da en el lenguaje cotidiano. Una característica fundamental de una distribución normal es que, cuando se grafica en un histograma, el resultado es una curva en forma de “campana”, como en el histograma generado con STATDISK que aquí se muestra. [Las principales características de la curva en forma de campana son **1.** el aumento de las frecuencias, las cuales alcanzan un punto máximo y luego disminuyen; y **2.** la simetría, donde la mitad izquierda de la gráfica es casi una imagen en espejo de la mitad derecha]. Este histograma corresponde a la distribución de frecuencias de la tabla 2-5, que incluye 1000 estaturas de mujeres elegidas al azar. Muchos procedimientos estadísticos requieren que los datos muestrales provengan de una población que tenga una distribución que no se desvíe mucho de la normalidad, y a menudo podemos usar un histograma para juzgar si se satisface tal requisito de una distribución normal.

Decimos que la distribución es *normal* porque es una curva en forma de campana.



## Uso de la tecnología

Los paquetes de cómputo estadísticos actuales son muy eficaces para generar gráficas impresionantes, incluyendo histogramas. En este libro a menudo hacemos referencia a STATDISK, Minitab, Excel, y la calculadora TI-83/84 Plus, que son tecnologías que sirven para generar histogramas. Las instrucciones detalladas varían desde lo extremadamente sencillo a lo muy complejo, por lo que a continuación haremos algunos co-

mentarios pertinentes. Para los procedimientos detallados, consulte los manuales que complementan este libro.

**STATDISK** genera histogramas con facilidad. Introduzca los datos en la ventana de datos de STATDISK, haga clic en **Data**, luego en **Histogram** y luego en el botón **Plot**. (Si usted prefiere determinar su propia anchura de clase y punto de partida, haga clic en el botón “User defined” antes de seleccionar Plot).

**MINITAB** genera histogramas con facilidad. Introduzca los datos en una columna, luego haga clic en **Graph** y luego en **Histogram**. Seleccione el histograma “Simple”. Introduzca la columna en la ventana “Graph variables” y seleccione **OK**.

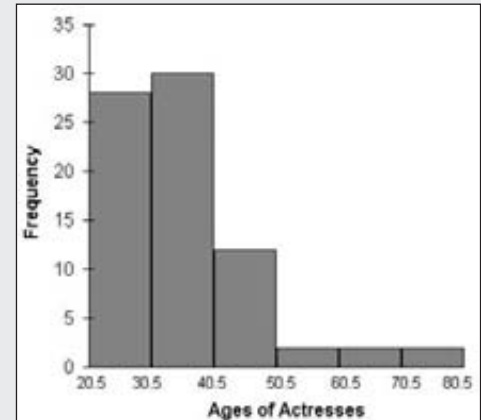
**TI-83/84 PLUS** Introduzca una lista de datos en L1. Seleccione la función **STAT PLOT** presionando [2nd] [Y=]. Presione [ENTER] y utilice las teclas de las flechas para poner P1 en la posición de en-

cendido y también elija la gráfica con barras. Presione [ZOOM] [9] para obtener un histograma con los valores predeterminados. (También puede utilizar su propia anchura de clase y fronteras de clase. Consulte el manual de TI-83/84 Plus que complementa este libro).

**EXCEL** puede generar histogramas como el que se presenta aquí, pero se vuelve algo difícil. Para generarlo con facilidad utilice el complemento DDXL que viene incluido en el CD de este libro. Después de

instalar DDXL dentro de Excel, haga clic en **DDXL**, seleccione **Charts and Plots** y haga clic en “function type” de **Histogram**. Haga clic en el icono del lápiz e introduzca el intervalo de celdas que contienen los datos, como A1:A500 para 500 valores en los renglones 1 a 500 de la columna A.

Excel



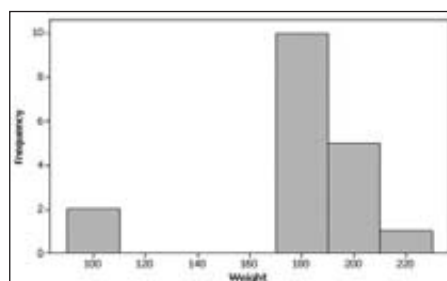
## 2-3 DESTREZAS Y CONCEPTOS BÁSICOS

### Conocimientos estadísticos y pensamiento crítico

- Histograma.** ¿Qué características importantes de los datos se entenderían mejor al examinar un histograma?
- Histograma y distribución de frecuencias.** Puesto que un histograma es básicamente una representación gráfica de los mismos datos de una distribución de frecuencias, ¿qué ventaja importante tiene un histograma sobre una distribución de frecuencias?
- Conjunto de datos pequeño.** Si un conjunto de datos es pequeño, por ejemplo de sólo cinco valores, ¿por qué no debemos tomarnos la molestia de construir un histograma?
- Distribución normal.** Después de examinar un histograma, ¿qué criterio se puede utilizar para determinar si los datos tienen una distribución aproximadamente normal? ¿Este criterio es totalmente objetivo o implica un juicio subjetivo?

*En los ejercicios 5 a 8, responda las preguntas con respecto al histograma que se generó con Minitab. El histograma representa los pesos (en libras) de los timoneles y los remeros de una carrera de embarcaciones entre Oxford y Cambridge. (Con base en datos de A Handbook of Small Data Sets, por D. J. Hand, Chapman y Hall).*

Histograma de Minitab





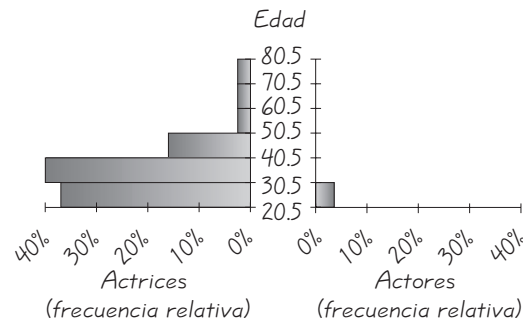
5. **Tamaño de la muestra.** ¿Cuántos miembros de la tripulación están incluidos en el histograma?
6. **Variación.** ¿Cuál es el peso mínimo posible? ¿Cuál es el peso máximo posible?
7. **Hueco.** ¿Cuál sería una explicación razonable para el gran hueco que hay entre la barra del extremo izquierdo y las otras barras?
8. **Anchura de clase.** ¿Cuál es la anchura de clase?
9. **Análisis del último dígito.** Remítase al ejercicio 17 de la sección 2-2, sobre los últimos dígitos de las estaturas de los estudiantes de estadística que se obtuvieron para un experimento de la clase. Utilice la distribución de frecuencias del ejercicio para construir un histograma. ¿Qué se puede concluir a partir de la distribución de los dígitos? En específico, ¿parece que las estaturas se reportaron o que realmente se midieron?
10. **Dado cargado.** Remítase al ejercicio 18 de la sección 2-2 sobre los resultados de 180 lanzamientos de un dado que el autor cargó. Utilice la distribución de frecuencias para construir el histograma correspondiente. ¿Cuál debería de ser la apariencia del histograma si el dado fuera perfectamente legal? ¿El histograma de la distribución de frecuencias dada parece diferir significativamente de un histograma obtenido de un dado legal?
11. **Cantidades de lluvia.** Remítase al ejercicio 19 de la sección 2-2 y utilice la distribución de frecuencias para construir un histograma. ¿Al parecer los datos tienen una distribución aproximadamente normal?
12. **Nicotina en cigarrillos.** Remítase al ejercicio 20 de la sección 2-2 y utilice la distribución de frecuencias para construir un histograma. ¿Al parecer los datos tienen una distribución aproximadamente normal?
13. **Valores del IMC.** Remítase al ejercicio 21 de la sección 2-2 y utilice la distribución de frecuencias para construir un histograma. ¿Al parecer los datos tienen una distribución aproximadamente normal?
14. **Datos del clima.** Remítase al ejercicio 22 de la sección 2-2 y utilice la distribución de frecuencias de las temperaturas mínimas reales para construir un histograma. ¿Al parecer los datos tienen una distribución aproximadamente normal?
15. **Pesos de monedas de un centavo.** Remítase al ejercicio 23 de la sección 2-2 y utilice la distribución de frecuencias con los pesos de las monedas de un centavo acuñadas antes de 1983. Construya el histograma correspondiente. ¿Los pesos parecen tener una distribución normal?
16. **Coca-Cola regular y Coca-Cola dietética.** Remítase al ejercicio 24 de la sección 2-2 y utilice las dos distribuciones de frecuencias relativas para construir los dos histogramas de frecuencias relativas correspondientes. Compare los resultados y determine si parece haber una diferencia significativa. Si hay una diferencia, ¿cómo se explicaría?
17. **Comparación de edades de actores y actrices.** Remítase la tabla 2-8 y utilice la distribución de frecuencias relativas de los mejores actores para construir un histograma de frecuencias relativas. Compare los resultados con la figura 2-3, que incluye el histograma de frecuencias relativas para las mejores actrices. ¿Al parecer los dos géneros ganan el premio Óscar a diferentes edades? (Véase también el ejercicio 18 de esta sección).

## 2-3 MÁS ALLÁ DE LO BÁSICO

18. **Histogramas de frecuencias relativas contiguos.** Cuando se utilizan histogramas para comparar dos conjuntos de datos, en ocasiones es difícil hacer comparaciones al cambiar



la vista de un histograma a otro. Un *histograma de frecuencias relativas contiguo* emplea un formato que facilita las comparaciones. En vez de las frecuencias, deberíamos utilizar las frecuencias relativas para que las comparaciones no se vean distorsionadas por muestras de tamaños diferentes. Complete los histogramas de frecuencias relativas contiguos que se indican a continuación, utilizando los datos de la tabla 2-8 de la sección 2-2. Luego utilice los resultados para comparar los dos conjuntos de datos.

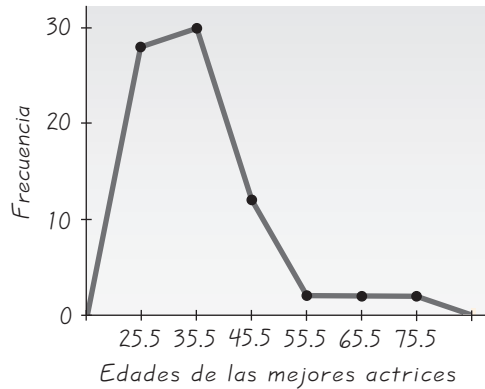


19. **Conjuntos grandes de datos.** Remítase al ejercicio 25 de la sección 2-2 y construya histogramas de frecuencias relativas contiguos para las cargas axiales de las latas de 0.0109 pulgadas de grosor y de las cargas axiales de las latas de 0.0111 pulgadas de grosor. (Los histogramas de frecuencias relativas contiguos se describen en el ejercicio 18). Compare los dos conjuntos de datos. ¿El grosor de las latas de aluminio afecta su resistencia, la cual se midió usando las cargas axiales?
20. **Interpretación de los efectos de los valores extremos.** Remítase al conjunto de datos 15 del apéndice B sobre las cargas axiales de las latas de aluminio de 0.0111 pulgadas de grosor. La carga de 504 lb es un *valor extremo* porque se encuentra muy alejado de los demás valores. Construya un histograma que incluya el valor de 504 lb, y luego construya otro histograma sin este valor. En ambos casos, inicie la primera clase en 200 lb y utilice una anchura de clase de 20 lb. Interprete los resultados planteando una generalización sobre el efecto que tendría un valor extremo en un histograma. (Véase el ejercicio 26 en la sección 2-2).

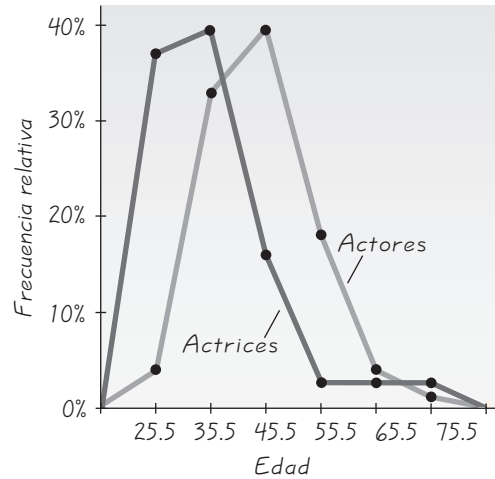
## 2-4 Gráficas estadísticas

**Concepto clave** En la sección 2-3 nos referimos a los histogramas y a los histogramas de frecuencias relativas como gráficas que muestran visualmente las distribuciones de conjuntos de datos. En esta sección presentamos otras gráficas de uso común en los análisis estadísticos, así como también gráficas que describen datos de formas innovadoras. Al igual que en la sección 2-3, el principal objetivo no es la creación de una gráfica, sino *entender* mejor un conjunto de datos usando de una gráfica adecuada, que revele de manera efectiva algunas características importantes. Nuestro mundo requiere de más personas con habilidades para construir gráficas que revelen de manera clara y efectiva las características importantes de los datos. Nuestro mundo también necesita gente con la capacidad de crear gráficas originales e innovadoras que muestren aspectos fundamentales de los datos.

Esta sección comienza con una descripción breve de las gráficas que generalmente se incluyen en los cursos de introducción a la estadística, como polígonos de frecuencias, ojivas, gráficas de puntos, gráficas de tallo y hojas, gráficas de



**Figura 2-4** Polígono de frecuencias



**Figura 2-5** Polígonos de frecuencias relativas

Pareto, gráficas circulares, diagramas de dispersión y gráficas de series de tiempo. Luego, estudiaremos algunas gráficas originales y creativas; ahora iniciaremos con los polígonos de frecuencias.

## Polígono de frecuencias

Un **polígono de frecuencias** utiliza segmentos lineales conectados a puntos que se localizan directamente por encima de los valores de las marcas de clase. Observe la figura 2-4, que presenta el polígono de frecuencias correspondiente a la tabla 2-2. Las alturas de los puntos corresponden a las frecuencias de clase; en tanto que los segmentos lineales se extienden hacia la derecha y hacia la izquierda, de manera que la gráfica inicia y termina sobre el eje horizontal.

Una variante del polígono de frecuencias básico es el polígono de frecuencias relativas, que coloca a estas últimas en la escala vertical. Al tratar de comparar dos conjuntos de datos, a menudo es muy útil graficar dos polígonos de frecuencias relativas sobre los mismos ejes. Consulte la figura 2-5, que ilustra polígonos de frecuencias relativas para las edades de las mejores actrices y los mejores actores, listadas en el Problema del capítulo. La figura 2-5 aclara visualmente que las actrices tienden a ser más jóvenes que sus contrapartes masculinos. La figura 2-5 logra algo que es realmente maravilloso: permite una comprensión de los datos que no sería posible mediante el examen visual de las listas de datos de la tabla 2-1. (Es como un buen maestro de poesía que revela el significado real de un poema). Por razones que no se explicarán aquí, parece que hay algún tipo de discriminación por género basado en la edad.

## Ojiva

Una **ojiva** es una gráfica lineal que representa frecuencias *acumulativas*, de la misma forma que la distribución de frecuencias acumulativas (véase la tabla 2-4 de la sección anterior) es una lista de éstas. La figura 2-6 es la ojiva correspondiente a la tabla 2-4. Observe que la ojiva utiliza fronteras de clase a lo largo de la escala horizontal, y que la gráfica comienza con la frontera inferior de la primera clase y termina con la frontera superior de la última clase. Las ojivas son útiles

- c) Si las ventas menores de siete automóviles por mes se consideran como un desempeño inaceptable, ¿cuál de los dos incisos, a) o b), es de mayor utilidad para identificar al grupo insatisfactorio de vendedores?
- 2-32 Kessler's Ice Cream Delight intenta mantener los 55 sabores diferentes de helado en existencia en cada una de sus sucursales. El director de investigación de mercado de la compañía sugiere que mantener un mejor registro para cada tienda es la clave para evitar quedarse sin existencias. Don Martin, director de operaciones, recolecta datos redondeados al medio galón más cercano sobre la cantidad diaria de helado de cada sabor que se vende. Nunca se venden más de 20 galones de un solo sabor al día.
- a) ¿La clasificación con respecto a los sabores es discreta o continua? ¿Abierta o cerrada?
- b) ¿La clasificación con respecto a la "cantidad de helado" es discreta o continua? ¿Abierta o cerrada?
- c) ¿Son los datos cuantitativos o cualitativos?
- d) ¿Qué le sugeriría a Don Martin para generar mejores datos para llevar a cabo una investigación de mercado?
- 2-33 Doug Atkinson es dueño y recogedor de boletos de un trasbordador que transporta pasajeros y automóviles desde Long Island hasta Connecticut. Doug tiene datos que indican el número de personas y de automóviles que han utilizado el trasbordador durante los dos últimos meses. Por ejemplo,
- 3 de julio   Número de pasajeros 173   Número de automóviles 32
- podría ser la información típica registrada durante un día. Doug ha establecido seis clases igualmente espaciadas para registrar el número diario de pasajeros y las marcas de clase son 84.5, 104.5, 124.5, 144.5, 164.5 y 184.5. Las seis clases igualmente espaciadas que construyó Doug para el número diario de automóviles tienen marcas de clase de 26.5, 34.5, 42.5, 50.5, 58.5 y 66.5. (Las marcas de clase son los puntos medios de los intervalos.)
- a) ¿Cuáles son los límites inferior y superior de las clases para el número de pasajeros?
- b) ¿Cuáles son los límites inferior y superior de las clases para el número de automóviles?

Soluciones a los ejercicios de autoevaluación

- |    |     |                     |      |      |       |       |       |
|----|-----|---------------------|------|------|-------|-------|-------|
| EA | 2-3 | Clase               | 1-6  | 7-12 | 13-18 | 19-24 | 25-30 |
|    |     | Frecuencia          | 4    | 8    | 4     | 3     | 1     |
|    |     | Frecuencia relativa | 0.20 | 0.40 | 0.20  | 0.15  | 0.05  |
- a) Suponiendo que la compañía abre 6 días a la semana, se ve que 80% de las órdenes se entregan en 3 semanas o menos.
- b) Se puede decir que sólo entre 20 y 60% de las entregas se hacen en 10 días o menos, de manera que la distribución no genera suficiente información para determinar si la meta se cumple.
- c) Una distribución de frecuencias relativas permite presentar frecuencias como fracciones o porcentajes.
- EA 2-4 La distribución no contiene todos los datos. El valor 500°C no aparece, tampoco los puntos entre 541 y 549°C, inclusive. Además, la distribución está cerrada a la derecha, lo que elimina todos los datos mayores de 600°C. Estas omisiones pueden explicar el hecho de que el número total de observaciones sea sólo 360, en lugar de 365 como podría esperarse del conjunto de datos coleccionados durante un año. (Nota: no es absolutamente necesario que la distribución esté abierta a la derecha, en especial si no se registraron valores mayores que 600°C. Sin embargo, para que esté completa, la distribución debe ser continua en el intervalo seleccionado, aun cuando no haya datos en algunos intervalos.) Por último, las clasificaciones no son mutuamente excluyentes. Los puntos 530°C y 580°C están contenidos en más de un intervalo. Al crear un conjunto continuo de clasificaciones, debe tenerse cuidado de evitar este error.

2.5 Representación gráfica de distribuciones de frecuencias

Identificación de los ejes horizontal y vertical

Las figuras 2-1 y 2-2 son un avance de lo que estudiaremos a continuación: cómo presentar las distribuciones de frecuencias de manera gráfica. Las gráficas proporcionan datos en un diagrama de dos dimensiones. En el eje *horizontal* podemos mostrar los valores de la variable (la característica que estamos midiendo), como la producción de alfombras en yardas. En el eje *vertical* señalamos las frecuencias de las clases mostradas en el eje horizontal. De esta forma, la altura de las barras de la figura

Función de las gráficas

2-1 mide el número de observaciones que hay en cada clase señalada en el eje horizontal. Las gráficas de distribuciones de frecuencias y de distribuciones de frecuencias relativas son útiles debido a que resaltan y aclaran los patrones que no se pueden distinguir fácilmente en las tablas. Atraen la atención del que las observa hacia los patrones existentes en los datos. Las gráficas también ayudan a resolver problemas relacionados a las distribuciones de frecuencias; nos permiten estimar algunos valores con sólo una mirada y proporcionan una verificación visual sobre la precisión de nuestras soluciones.

## Histogramas

Descripción de los histogramas

Las figuras 2-1 y 2-2 son dos ejemplos de histogramas. Un *histograma* consiste en una serie de rectángulos, cuyo ancho es proporcional al rango de los valores que se encuentran dentro de una clase, y cuya altura es proporcional al número de elementos que caen dentro de la clase. Si las clases empleadas en la distribución de frecuencias son del mismo ancho, entonces las barras verticales del histograma también tienen el mismo ancho. La altura de la barra correspondiente a cada clase representa el número de observaciones de la clase. Como consecuencia, el área contenida en cada rectángulo (base por altura) ocupa un porcentaje del área total de todos los rectángulos la cual es igual a la frecuencia absoluta de esa clase correspondiente respecto a todas las observaciones hechas.

Función de un histograma de frecuencias relativas

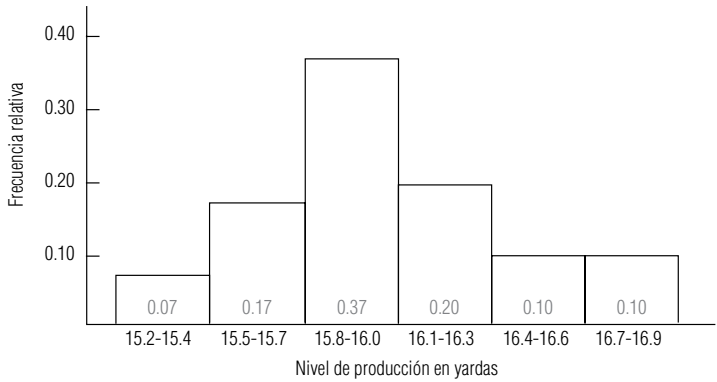
Un histograma que utiliza las frecuencias relativas de los datos puntuales de cada una de las clases, en lugar de usar el número real de puntos, se conoce como *histograma de frecuencias relativas*. Este tipo de histograma tiene la misma forma que un histograma de frecuencias absolutas construido a partir del mismo conjunto de datos. Esto es así debido a que en ambos, el tamaño relativo de cada rectángulo es la frecuencia de esa clase comparada con el número total de observaciones.

Recuerde que la frecuencia relativa de cualquier clase es el número de observaciones que entran en la clase, dividido entre el número total de observaciones hechas. La suma de todas las frecuencias relativas de cualquier conjunto de datos debe ser igual a 1.0. Con esto en mente, podemos convertir el histograma de la figura 2-1 en un histograma de frecuencias relativas como el presentado en la figura 2-7. Observe que la única diferencia entre éstos es el lado izquierdo de la escala del eje vertical. Mientras que la escala vertical del histograma de la figura 2-1 representa el número *absoluto* de observaciones de cada clase, la escala del histograma de la figura 2-7 es el número de observaciones de cada clase tomadas como una *fracción* del número total de observaciones.

Ventajas del histograma de frecuencias relativas

Poder presentar los datos en términos de la frecuencia relativa de las observaciones, más que en términos de la frecuencia absoluta, es de gran utilidad, ya que mientras los números absolutos pueden sufrir cambios (si probamos más telares, por ejemplo), la relación entre las clases permanece estable. El 20% de todos los telares puede entrar en la clase “16.1-16.3 yardas”, ya sea que probemos 30 o 300 telares. Resulta fácil comparar los datos de muestras de diferentes tamaños cuando utilizamos histogramas de frecuencias relativas.

**FIGURA 2-7**  
Distribución de frecuencias relativas de los niveles de producción de una muestra de 30 telares para alfombra utilizando intervalos de clase de 0.3 yardas



## Polígonos de frecuencias

Utilice los puntos medios en el eje horizontal

Añada dos clases

Conversión de un polígono de frecuencias en un histograma

Construcción de un polígono de frecuencias relativas

Ventajas del histograma

Aunque se utilizan menos, los polígonos de frecuencias son otra forma de representar gráficamente distribuciones tanto de frecuencias como de frecuencias relativas. Para construir un polígono de frecuencias señalamos éstas en el eje vertical y los valores de la variable que estamos midiendo en el eje horizontal, del mismo modo en que se hizo con el histograma. A continuación, graficamos cada frecuencia de clase trazando un punto sobre su punto medio y conectamos los puntos sucesivos resultantes con una línea recta para formar un polígono (una figura con muchos lados).

La figura 2-8 representa un polígono de frecuencias construido a partir de los datos de la tabla 2-14. Si compara esta figura con la 2-1, notará que se han agregado dos clases, una en *cada extremo* de la escala de valores observados. Éstas contienen cero observaciones, pero permiten que el polígono llegue al eje horizontal en ambos extremos de la distribución.

¿De qué manera podemos convertir un polígono de frecuencias en un histograma? Un polígono de frecuencias es sólo una línea que conecta los puntos medios de todas las barras de un histograma. Por consiguiente, podemos reproducir el histograma mediante el trazado de líneas verticales desde los límites de clase (señalados en el eje horizontal) y, luego, conectando esas líneas con rectas horizontales a la altura de los puntos medios del polígono. En la figura 2-9 hicimos esto con líneas punteadas.

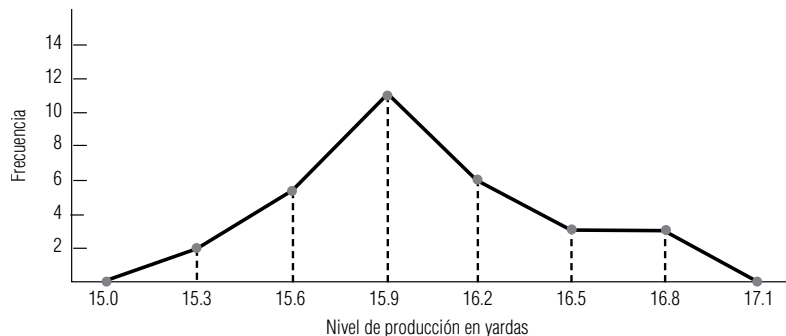
Un polígono de frecuencias que utiliza frecuencias relativas de datos puntuales en cada una de las clases, en lugar del número real de puntos, se conoce como *polígono de frecuencias relativas*. Este polígono tiene la misma forma que el polígono de frecuencias construido a partir del mismo conjunto de datos, pero con una escala diferente en los valores del eje vertical. En lugar del número absoluto de observaciones, la escala representa el número de observaciones de cada clase expresadas como una fracción del total de observaciones.

Los histogramas y los polígonos de frecuencias son similares. ¿Por qué necesitamos ambos? Las ventajas de los histogramas son:

1. Los rectángulos muestran cada clase de la distribución por separado.
2. El área de cada rectángulo, en relación con el resto, muestra la proporción del número total de observaciones que se encuentran en esa clase.

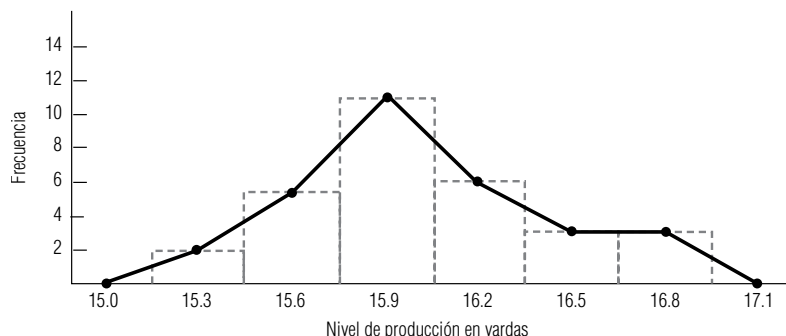
**FIGURA 2-8**

Polígono de frecuencias del nivel de producción de una muestra de 30 telares para alfombra utilizando intervalos de clase de 0.3 yardas



**FIGURA 2-9**

Histograma trazado a partir de los puntos del polígono de frecuencias de la figura 2-8



Los polígonos, por su parte, también poseen ciertas ventajas.

- 1. El polígono de frecuencias es más sencillo que su histograma correspondiente.
- 2. Bosqueja con más claridad un perfil del patrón de los datos.
- 3. El polígono se vuelve cada vez más suave y parecido a una curva conforme aumentamos el número de clases y el número de observaciones.

Un polígono como el que acabamos de describir, suavizado mediante el aumento de clases y de datos puntuales, se conoce como *curva de frecuencias*. En la figura 2-10, hemos utilizado el ejemplo de los telares para alfombra, pero en esta ocasión aumentamos el número de observaciones a 300 y el número de clases a 10. Note que conectamos los puntos con líneas curvas para tener una aproximación de la manera en que se vería el polígono si tuviéramos un gran número de datos puntuales e intervalos de clase muy pequeños.

Ojivas

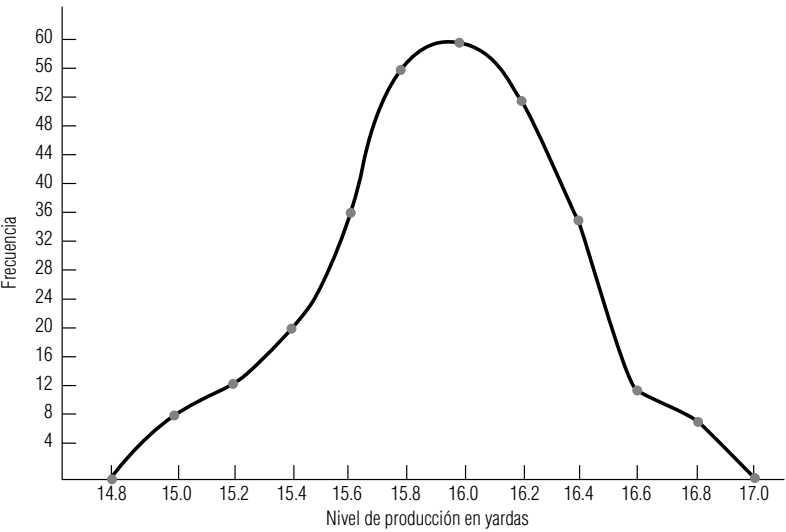
Una *distribución de frecuencias acumuladas* nos permite ver cuántas observaciones están por encima de ciertos valores, en lugar de hacer un mero registro del número de elementos que hay dentro de los intervalos. Por ejemplo, si deseamos saber cuántos telares tejen menos de 17.0 yardas, podemos utilizar una tabla que registre las frecuencias acumuladas “menores que” de nuestra muestra, como se presenta en la tabla 2-15.

La gráfica de una distribución de frecuencias acumuladas se conoce como *ojiva*. En la figura 2-11 se muestra la ojiva de la distribución de frecuencias de la tabla 2-15. Los puntos representados en la gráfica indican el número de telares que tienen una producción menor que el número de yardas mostrado en el eje horizontal. Note que el límite inferior de las clases de la tabla se convierte en el límite superior de la distribución acumulada de la ojiva.

Tabla 2-15	Clase	Frecuencia acumulada
Distribución de frecuencias “menores que” acumulada de los niveles de producción de una muestra de 30 telares para alfombra	Menos que 15.2	0
	Menos que 15.5	2
	Menos que 15.8	7
	Menos que 16.1	18
	Menos que 16.4	24
	Menos que 16.7	27
	Menos que 17.0	30

FIGURA 2-10

Curva de frecuencias de los niveles de producción de una muestra de 300 telares para alfombra utilizando intervalos de 0.2 yardas



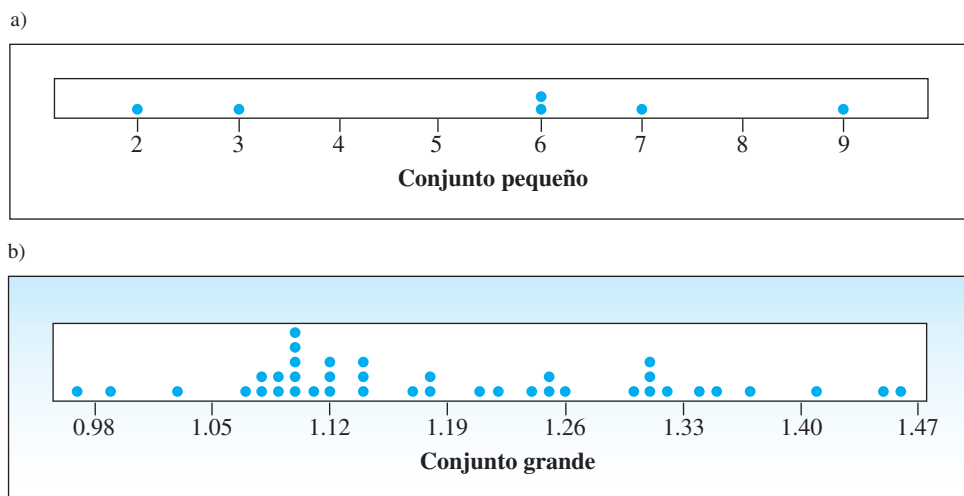
## Gráficas de puntos

Muchos conjuntos de datos cuantitativos están formados de números que no se pueden separar fácilmente en categorías o intervalos. Entonces se hace necesaria una forma diferente de graficar este tipo de datos.

La gráfica más sencilla para datos cuantitativos es la **gráfica de puntos**. Para un conjunto pequeño de mediciones, por ejemplo el conjunto 2, 6, 9, 3, 7, 6, se puede simplemente graficar las mediciones como puntos en un eje horizontal. Esta gráfica de puntos, generada por *MINITAB*, se muestra en la figura 1.9a). Para un conjunto grande de datos, como el de la figura 1.9b), la gráfica de puntos puede ser nada informativa y tediosa para interpretarse.

**FIGURA 1.9**

Gráficas de puntos para conjuntos pequeños y grandes de datos



## Gráficas de tallo y hoja

Otra forma sencilla de exhibir la distribución de un conjunto de datos cuantitativos es la **gráfica de tallo y hoja**. Esta gráfica presenta una exhibición gráfica de los datos usando los valores numéricos reales de cada punto de datos.

MI

### ENTRENADOR PERSONAL

#### ¿Cómo construyo una gráfica de tallo y hoja?

1. Divida cada segmento en dos partes: el **tallo** y las **hojas**.
2. Ponga en lista los tallos en una columna, con una línea vertical a su derecha.
3. Para cada medición, registre la parte de hoja en el mismo renglón como su tallo correspondiente.
4. Ordene las hojas de menor a mayor en cada tallo.
5. Dé una clave a su codificación de tallo y hoja para que el lector pueda recrear las mediciones reales si es necesario.

#### EJEMPLO

1.7

La tabla 1.7 es una lista de precios (en dólares) de 19 marcas de zapatos deportivos. Construya una gráfica de tallo y hoja para mostrar la distribución de los datos.



**TABLA 1.7**

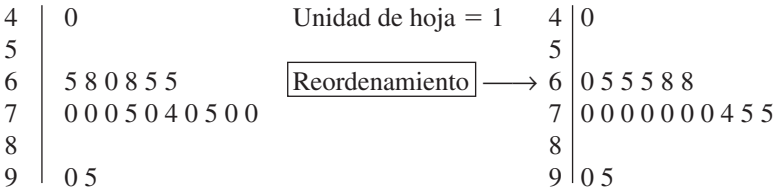
**Precios de zapatos deportivos**

90	70	70	70	75	70
65	68	60	74	70	95
75	70	68	65	40	65
70					

**Solución** Para crear el tallo y hoja, se puede dividir cada observación entre las unidades y las decenas. El número a la izquierda es el tallo; el de la derecha es la hoja. Entonces, para los zapatos que cuestan \$65, el tallo es 6 y la hoja es 5. Los tallos, que van de 4 a 9, aparecen en la figura 1.10, junto con las hojas para cada una de las 19 mediciones. Si indicamos que la unidad de hoja es 1, el lector verá que el tallo y hoja 6 y 8, por ejemplo, representan el número 68 registrado al dólar más cercano.

**FIGURA 1.10**

Gráfica de tallo y hoja para los datos de la tabla 1.7



**MI CONSEJO**  
tallo | hoja

A veces las opciones de tallo disponibles resultan en una gráfica que contiene muy pocos tallos y un gran número de hojas dentro de cada tallo. En esta situación, se pueden prolongar los tallos al dividir cada uno en varias líneas, dependiendo de los valores de hojas que se les asignen. Por lo general los tallos se dividen en una de dos formas:

- En dos líneas, con las hojas 0-4 en la primera línea y las hojas 5-9 en la segunda línea
- En cinco líneas, con las hojas 0-1, 2-3, 4-5, 6-7 y 8-9 en las cinco líneas, respectivamente

**EJEMPLO**

1.8

Los datos de la tabla 1.8 son los pesos de 30 bebés de gestación completa al momento de nacer, nacidos en un hospital metropolitano y registrados al décimo de libra más cercano.<sup>6</sup> Construya una gráfica de tallo y hoja para mostrar la distribución de los datos.

**TABLA 1.8**

**Pesos de 30 bebés de gestación completa al momento de nacer**

7.2	7.8	6.8	6.2	8.2
8.0	8.2	5.6	8.6	7.1
8.2	7.7	7.5	7.2	7.7
5.8	6.8	6.8	8.5	7.5
6.1	7.9	9.4	9.0	7.8
8.5	9.0	7.7	6.7	7.7

**Solución** Los datos, aun cuando están registrados a una precisión de sólo un lugar decimal, son mediciones de la variable continua  $x =$  peso, que puede tomar cualquier valor positivo. Al examinar la tabla 1.8, se puede ver rápidamente que los pesos más alto y más bajo son 9.4 y 5.6, respectivamente. Pero, ¿cómo están distribuidos los pesos restantes?

# 2

Johnson, edición 11

## Análisis descriptivo y presentación de datos de una variable



©2010 Alys Tomlinson/Jupiterimages



©2010 Chris Whitehead/Jupiterimages

### PRESENTACIÓN GRÁFICA DE DATOS

2.1 Gráficas, diagramas de Pareto y diagramas de tallo y hojas

*Una **imagen** vale más que mil palabras.*

2.2 Distribuciones de frecuencia e histogramas

*Métodos **gráficos** para conjuntos de datos más grandes.*

### ESTADÍSTICA DESCRIPTIVA NUMÉRICA

2.3 Medidas de tendencia central

*Media, mediana, moda y medio rango son valores **promedio**.*

2.4 Medidas de dispersión

*Cómo medir la **cantidad de dispersión** en un conjunto de datos.*

2.5 Medidas de posición

*Cómo **comparar** un valor de datos con el conjunto de datos.*

2.6 Interpretación y comprensión de la desviación estándar

*La longitud de una **vara de medir estandarizada**.*

2.7 El arte del engaño estadístico

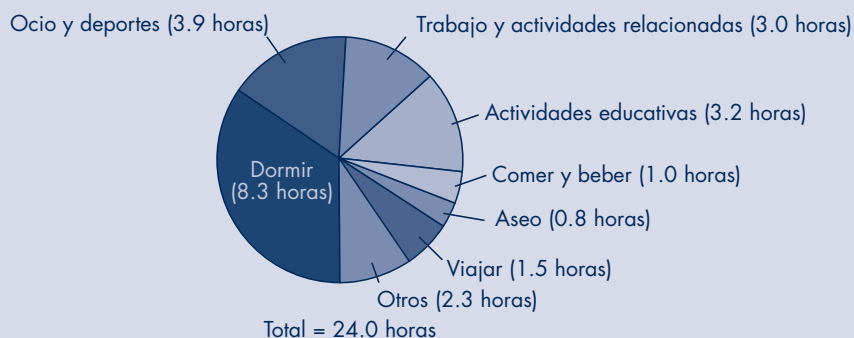
*Gráficas “**truculentas**” e **información insuficiente** confunden.*

## 2.1 Gráficas, diagramas de Pareto y diagramas de tallo y hojas

### Estudiantes: Aquí los observan

Considera toda la información en la gráfica, específicamente llamada gráfica de pastel o gráfica de círculo. ¿Tú día se divide en las categorías que se muestran en la siguiente página? ¿O tienes una o dos categorías adicionales? ¿Tal vez menos categorías? Ahora considera el tiempo otorgado

#### Uso de tiempo en un día promedio para estudiantes universitarios de tiempo completo



**NOTA:** Los datos incluyen individuos, con edades de 15 a 49 años, inscritos de tiempo completo en una universidad. Los datos incluyen fines de semana no festivos y son promedios para 2003-2007.

**Fuente:** Bureau of Labor Statistics

**PTI** ATUS es un sondeo continuo de la administración federal acerca del uso del tiempo en Estados Unidos, patrocinado por la Bureau of Labor Statistics y realizada por la U.S. Census Bureau

para cada actividad, en promedio, ¿cómo se compara la cantidad de tiempo que tú empleas? Quizá tú tienes categorías completamente diferentes. ¿Deseas tener las 8.3 horas de sueño, en promedio? ¡Los autores sí!

¿Puedes imaginar toda esta información escrita en oraciones? Las presentaciones gráficas verdaderamente pueden valer mil palabras. Esta gráfica de pastel resume la información “Uso del tiempo” de la Encuesta de Uso de Tiempo Estadounidense 2003-2007 (ATUS, por sus siglas en inglés) de más de 50 000 estadounidenses. Dado que se trata de un sondeo transversal, fíjate que esta gráfica sólo incluye a los estudiantes universitarios de tiempo completo que participaron.

Ahora que conoces la fuente y ves el tamaño global de la muestra, puedes sentir que dichos datos representan una imagen relativamente precisa de un día de un estudiante universitario. Tal vez quieras observar más de cerca alguna de las categorías. ¿Tienes preguntas acerca del promedio de 0.8 horas por día en aseo? ¿Crees que pueda haber una diferencia de género? Te hace pensar, ¿no es así?

**PTI** No hay una respuesta correcta exclusiva cuando construyes una presentación gráfica. El juicio del analista y las circunstancias que rodean el problema tienen importantes papeles en el desarrollo de la gráfica.

Como se demuestra con la gráfica de la página 32, una de las formas más útiles para familiarizarse con la información es usar una técnica de análisis inicial para explorar los datos que resultarán en una representación pictórica de los mismos. La presentación revelará visualmente patrones de comportamiento de la variable a estudiar. Existen varias formas gráficas (visuales) para describir la información. El tipo de datos y la idea a presentar determinan cuál método usar.

Datos cualitativos

**Gráficas de pastel (gráficas circulares) y gráficas de barras** Gráficas que se usan para resumir **datos cualitativos**, atributos o categóricos. Las gráficas de pastel (gráficas circulares) muestran la cantidad de datos que pertenecen a cada categoría como una parte proporcional de un círculo. Las gráficas de barras muestran la cantidad de datos que pertenecen a cada categoría como un área rectangular de tamaño proporcional.

EJEMPLO 2.1  
GRAFICACIÓN DE DATOS CUALITATIVOS

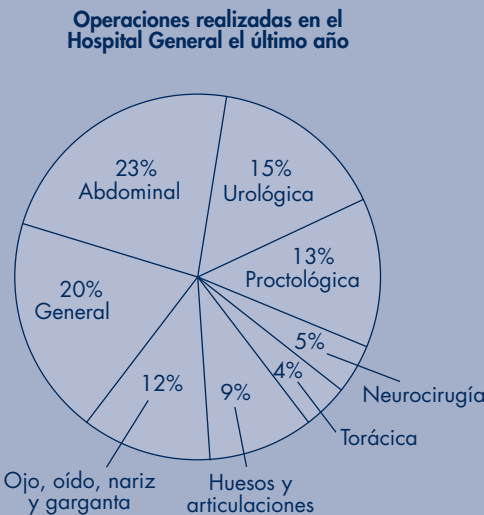
La tabla 2.1 presenta el número de casos de cada tipo de operación realizada en el Hospital General el último año.

**TABLA 2.1** Operaciones realizadas en el Hospital General el último año [TA02-01]

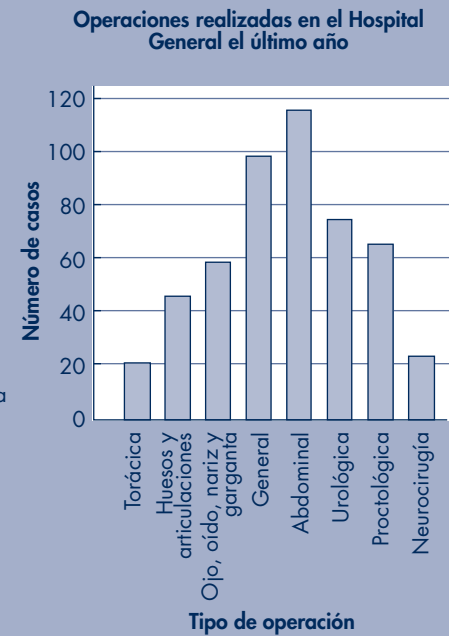
Tipo de operación	Número de casos
Torácica	20
Huesos y articulaciones	45
Ojo, oído, nariz y garganta	58
General	98
Abdominal	115
Urológica	74
Proctológica	65
Neurocirugía	23
Total	498

Los datos en la tabla 2.1 se muestran en una gráfica de pastel en la figura 2.1, donde cada tipo de operación se representa mediante una proporción relativa de un círculo, que se encuentra al dividir el número de casos por el tamaño muestral total, a saber, 498. Las proporciones se reportan entonces como porcentajes (por ejemplo, 25% es 1/4 del círculo). La figura 2.2 muestra los mismos datos de “tipo de operación”, pero en forma de una gráfica de barras. Las gráficas de barras de datos de atributo deben dibujarse con un espacio entre barras de igual ancho.

**FIGURA 2.1**  
Gráfica de pastel



**FIGURA 2.2**  
Gráfica de barras



**PTI** Todas las representaciones gráficas necesitan explicarse completamente a sí mismas. Esto incluye una descripción, título significativo e identificación adecuada de las cantidades y variables involucradas.

INSTRUCCIONES DE TECNOLOGÍA:  
GRÁFICA DE PASTEL

**MINITAB**

Escribe las categorías en C1 y las frecuencias correspondientes en C2; después continúa con:

- Elige: **Graph > Pie Chart . . .**
- Selecciona: **Chart Values from a table**
- Escribe: **Variable categórica: C1 Variables resumen: C2**
- Selecciona: **Labels > Title/Footnotes** Escribe: **Título: tu título**
- Selecciona: **Etiquetas deseadas > Select desired labels > OK > OK**

**Excel**

Escribe las categorías en la columna A y las frecuencias correspondientes en la columna B; activa ambas columnas de datos al resaltar y seleccionar los nombres de columna y las celdas de datos, después continúa con:

- Elige: **Insert > Pie > 1st picture** (usualmente)
- Elige: **Chart Layouts—Layout 1**
- Escribe: **Chart title: Tu título**

Para editar la gráfica de pastel:

Haz clic en: Cualquier parte para limpiar la gráfica (usa las manijas para el tamaño)  
Cualquier celda en la categoría o columna de frecuencia y escribe diferentes nombres o cantidades > ENTER

### TI-83/84 Plus

Escribe las frecuencias para las diversas categorías en L1; después continúa con:

Elige: PRGM > EXEC > CIRCLE\*  
Escribe: LIST: L1 > ENTER  
DATA DISPLAYED?: 1:PERCENTAGES  
OR  
2:DATA



\* El programa "CIRCLE" de la TI-83/84 Plus y otros programas están disponibles para descarga a través de [cengagebrain.com](http://cengagebrain.com). Los programas de la TI-83/84 Plus y los archivos de datos pueden estar en formato zip o comprimido. Si es así, guarda los archivos y descomprímelos usando una utilidad zip. Descarga los programas a tu calculadora usando el software TI-Graph Link.

Cuando la gráfica de barras se presenta en la forma de un *diagrama de Pareto*, presenta información adicional y muy útil.

**Diagrama de Pareto** Gráfica de barra con las barras ordenadas de la categoría más numerosa a la categoría menos numerosa. Incluye una gráfica de línea que muestra los porcentajes acumulados y conteos de las barras.

El diagrama de Pareto es popular en aplicaciones de control de calidad. Un diagrama de Pareto de tipos de defecto mostrará aquellos que tengan el mayor efecto sobre la tasa de defectos en orden de efecto. Entonces es fácil ver cuáles defectos deben observarse para reducir de manera más efectiva la tasa de defectos.

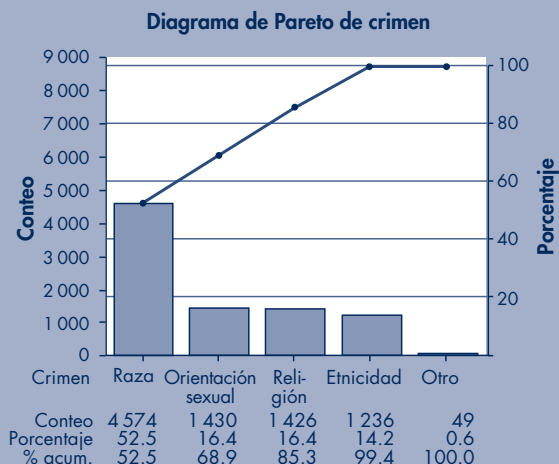
## EJEMPLO 2.2



### DIAGRAMA DE PARETO DE CRÍMENES DE ODIO

El FBI reportó el número de crímenes de odio por categoría para 2003 (<http://www.fbi.gov/>). El diagrama de Pareto de la figura 2.3 muestra los 8 715 crímenes de odio por categoría, sus porcentajes y porcentajes acumulados.

**FIGURA 2.3**  
Diagrama de Pareto



## INSTRUCCIONES DE TECNOLOGÍA: DIAGRAMA DE PARETO

### MINITAB

Escribe las categorías en C1 y las frecuencias correspondientes en C2; después continúa con:

Elige: **Start Chart > Quality Tools > Pareto**  
 Selecciona: **Chart defects table**  
 Escribe: Datos de defectos o atributo en: C1  
 Frecuencias en: C2  
 Selecciona: **Options**  
 Escribe: Title: **tu título > OK > OK**

### Excel

Escribe las categorías en la columna A y las frecuencias correspondientes en la columna B (los encabezados de columna son opcionales); después continúa con:  
 Primero ordena la tabla:

Activa ambas columnas de la distribución

Elige: **Data > AZ / ZA Short**  
 Selecciona: Story by: **frequency column**  
 Order: **Largest to Smallest > OK**  
 Elige: **Insert > Column > 1st picture** (usualmente)  
 Elige: Chart Layouts—Layout 9  
 Escribe: Título gráfica: **tu título**  
 Título eje categoría (x): **título para eje x**  
 Título eje valor (y): **título para eje y**

Para editar el diagrama de Pareto:

Haz clic en: Cualquier parte para limpiar la gráfica (usa las manijas para el tamaño)  
 Cualquier nombre de título para cambiarlo  
 Cualquier celda en la columna de categoría y escribe un nombre > **Enter**

Excel no incluye la gráfica de línea.

### TI-83/84 Plus

Escribe las categorías numeradas en L1 y las frecuencias correspondientes en L2; después continúa con:

Elige: **PRGM > EXEC > PARETO\***  
 Escribe: **LIST: L2 > ENTER**  
 Ymax: **al menos la suma de las frecuencias > ENTER**  
 Yscl: **incremento para eje y > ENTER**



\*El programa "PARETO" es uno de muchos programas que están disponibles para descargar. Véase la página 35 para instrucciones específicas.

## Datos cuantitativos

Una de las principales razones para construir una gráfica de **datos cuantitativos** es mostrar su *distribución*.

**Distribución** Patrón de variabilidad que muestran los datos de una variable. La distribución muestra la frecuencia de cada valor de la variable.

Una de las gráficas más simples usadas para mostrar una distribución es la *gráfica de puntos*.

**Gráfica de puntos** Describe los datos de una muestra al representar cada valor de datos con un punto colocado a lo largo de una escala. Esta escala puede ser horizontal o vertical. La frecuencia de los valores se representa a lo largo de la otra escala.

EJEMPLO 2.3

GRÁFICA DE PUNTOS DE CALIFICACIONES DE EXAMEN

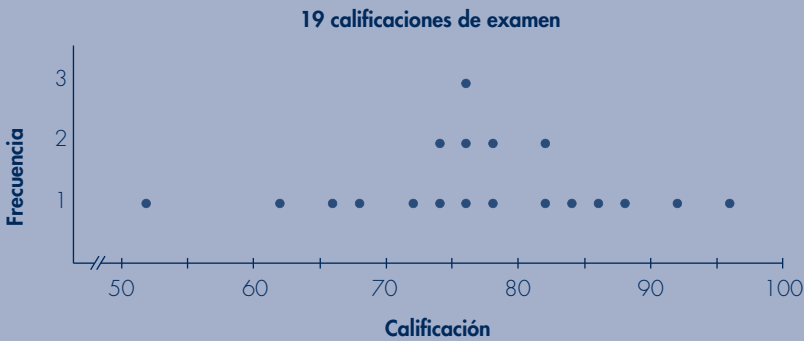
La tabla 2.2 proporciona una muestra de 19 calificaciones de examen seleccionadas al azar de una clase grande.

**TABLA 2.2**  
Muestra de 19 calificaciones de examen [TA02-02]

76	74	82	96	66	76	78	72	52	68
86	84	62	76	78	92	82	74	88	

La figura 2.4 es una gráfica de puntos de las 19 calificaciones de examen.

**FIGURA 2.4**  
Gráfica de puntos



Nota cómo los datos de la figura 2.4 están “apiñados” cerca del centro y más “dispersos” cerca de los extremos.

La gráfica de puntos es una técnica conveniente que se usa cuando uno comienza a analizar los datos. Resulta en una imagen de los datos que los ordena numéricamente. (*Ordenar* los datos es hacer una lista de los mismos en una clasificación organizada de acuerdo con el valor numérico.)

INSTRUCCIONES DE TECNOLOGÍA:  
GRÁFICA DE PUNTOS

MINITAB

Escribe los datos en C1; después continúa con:

Elige: **Graph > Dotplot . . . > One Y, Simple > OK**  
Escribe: **Graph Variables: C1 > OK**

Excel

La gráfica de puntos no está disponible, pero puedes hacer el paso inicial de clasificar los datos. Escribe los datos en la columna A y activa la columna de datos; después continúa con:

Elige: **Data > AZ ↓ (Sort)**  
Use los datos ordenados para terminar de construir la gráfica de puntos.



TI-83/84 Plus

Escribe los datos en L1; después continúa con:

Elige: PRGM > EXEC > DOTPLOT\*  
Escribe: LIST: L1 > ENTER  
Xmin: cuando mucho el valor x más bajo  
Xmax: al menos el valor x más alto  
Xscl: 0 o incremento  
Ymax: al menos la frecuencia más alta



\*El programa "DOTPLOT" es uno de muchos programas que están disponibles para descargar. Véase la página 35 para instrucciones específicas.

En años recientes se ha vuelto popular una técnica conocida como *presentación de tallo y hojas* para resumir datos numéricos. Es una combinación de una técnica gráfica y una técnica de ordenación. Dichas presentaciones son simples de crear y usar y son bastante adecuadas para aplicaciones de cómputo.

**Presentación de tallo y hojas** Presenta los datos de una muestra con los dígitos reales que constituyen los valores de datos. Cada valor numérico se divide en dos partes: el (los) dígito(s) inicial(es) es (son) el tallo y los dígitos posteriores son las hojas. Los tallos se ubican a lo largo del eje principal y para cada valor de datos se ubica una hoja de modo que muestre la distribución de los datos.

EJEMPLO 2.4

FIGURA 2.5A  
Presentación sin terminar de tallo y hojas

19 calificaciones de examen

5	2
6	6 8 2
7	6 4 6 8 2 6 8 4
8	2 6 4 2 8
9	6 2

FIGURA 2.5B  
Presentación final de tallo y hojas

19 calificaciones de examen

5	2
6	2 6 8
7	2 4 4 6 6 6 8 8
8	2 2 4 6 8
9	2 6

CONSTRUCCIÓN DE UNA PRESENTACIÓN DE TALLO Y HOJAS

Ahora construye una presentación de tallo y hojas para las 19 calificaciones de examen que se proporcionan en la tabla 2.2 de la página 37.

En un vistazo rápido podrás ver que hay calificaciones en los 50, 60, 70, 80 y 90. Usa el primer dígito de cada calificación como el tallo y el segundo dígito como la hoja. Por lo general, la presentación se construye verticalmente. Traza una línea vertical y coloca los tallos, en orden, a la izquierda de la línea.

5  
6  
7  
8  
9

A continuación coloca cada hoja sobre su tallo. Esto se hace al colocar el dígito posterior a la derecha de la línea vertical opuesta a su correspondiente dígito inicial. El primer valor de datos es 76; 7 es el tallo y 6 es la hoja. Por tanto, coloca un 6 opuesto al tallo 7:

7 | 6

El siguiente valor de datos es 74, de modo que una hoja 4 se coloca en el tallo 7 junto al 6.

7 | 6 4

**FIGURA 2.6**  
Presentación de tallo y hojas

19 calificaciones de examen

(50–54) 5	2
(55–59) 5	
(60–64) 6	2
(65–69) 6	6 8
(70–74) 7	2 4 4
(75–79) 7	6 6 6 8 8
(80–84) 8	2 2 4
(85–89) 8	6 8
(90–94) 9	2
(95–99) 9	6

El siguiente valor de datos es 82, de modo que una hoja 2 se coloca en el tallo 8.

7 | 6 4  
8 | 2

Continúa hasta que cada una de las otras 16 hojas se coloque en la presentación. La figura 2.5A muestra la presentación resultante en tallo y hojas; la figura 2.5B muestra la presentación completa de tallo y hojas después de ordenar las hojas.

A partir de la figura 2.5B, puedes ver que las calificaciones se centran alrededor de los 70. En este caso todas las calificaciones con los mismos dígitos de decenas se colocaron sobre la misma rama, pero esto puede no ser siempre deseable. Supón que reconstruyes la presentación; esta vez, en lugar de agrupar 10 posibles valores en cada tallo, agrupas los valores de modo que sólo 5 posibles valores puedan caer en cada tallo, como se muestra en la figura 2.6. ¿Observas alguna diferencia en la apariencia de la figura 2.6?, la forma general es aproximadamente simétrica en torno al alto de los 70. La información está un poco más refinada, pero básicamente se ve la misma distribución.

INSTRUCCIONES DE TECNOLOGÍA:  
PRESENTACIONES DE TALLO Y HOJAS

MINITAB

Escribe los datos en C1; después continúa con:

Elige: **Graph > Stem-and-Leaf ...**  
Escribe: **Graph variables: C1**  
**Increment: ancho de tallo (opcional) > OK**

Excel

Escribe los datos en la columna A; después continúa con:

Elige: **Add-Ins > Data Analysis Plus\* > Stem and Leaf Display > OK**  
Escribe: **Input Range: (A2:A6 o selecciona celdas)**  
**Increment: Stem Increment**



\*Data Analysis Plus es una colección de macros estadísticas para Excel y uno de los muchos programas disponibles para descargar a través de [cengagebrain.com](http://cengagebrain.com).

TI-83/84 Plus

Escribe los datos en L1; después continúa con:

Elige: **STAT > EDIT > 2:SortA(**  
Escribe: **L1**

Usa los datos ordenados para terminar de construir a mano el diagrama de tallo y hojas.

Es bastante usual que muchas variables presenten una distribución que esté concentrada (ajustada) en torno a un valor central y después en alguna forma dispersa en una o ambas direcciones. Con frecuencia, una presentación gráfica revela algo que el analista puede o no haber anticipado. El ejemplo 2.5 demuestra lo que en general ocurre cuando dos poblaciones se muestrean juntas.

EJEMPLO 2.5



DISTRIBUCIONES TRASLAPADAS

Se selecciona una muestra aleatoria de 50 estudiantes universitarios. Sus pesos se obtienen a partir de sus registros médicos. Los datos resultantes se muestran en la tabla 2.3.

Observa que los pesos varían de 98 a 215 libras. Agrupa los pesos en tallos de 10 unidades, usando los dígitos de centenas y decenas como tallos y los dígitos de unidades como la hoja (véase la figura 2.7). Las hojas se ordenaron numéricamente.

Una inspección cercana de la figura 2.7 sugiere que pueden estar involucradas dos distribuciones traslapadas. Esto es exactamente lo que se tiene: una distribución de pesos de mujeres y una distribución de pesos de hombres. La figura 2.8 muestra una presentación de tallo y hojas “espalda con espalda” de este conjunto de datos y hace obvio que están involucradas dos distribuciones distintas.

TABLA 2.3  
Pesos de 50 estudiantes universitarios [TA02-03]

Estudiante	1	2	3	4	5	6	7	8	9	10
Hombre/Mujer	M	H	M	H	H	M	M	H	H	M
Peso	98	150	108	158	162	112	118	167	170	120
Estudiante	11	12	13	14	15	16	17	18	19	20
Hombre/Mujer	H	H	H	M	M	H	M	H	H	M
Peso	177	186	191	128	135	195	137	205	190	120
Estudiante	21	22	23	24	25	26	27	28	29	30
Hombre/Mujer	H	H	M	H	M	M	H	H	H	H
Peso	188	176	118	168	115	115	162	157	154	148
Estudiante	31	32	33	34	35	36	37	38	39	40
Hombre/Mujer	M	H	H	M	H	M	H	M	H	H
Peso	101	143	145	108	155	110	154	116	161	165
Estudiante	41	42	43	44	45	46	47	48	49	50
Hombre/Mujer	M	H	M	H	H	M	M	H	H	H
Peso	142	184	120	170	195	132	129	215	176	183

FIGURA 2.7  
Presentación de tallo y hojas

Pesos de 50 estudiantes universitarios (lb)	
N = 50	Unidad hoja = 1.0
9	8
10	1 8 8
11	0 2 5 5 6 8 8
12	0 0 0 8 9
13	2 5 7
14	2 3 5 8
15	0 4 4 5 7 8
16	1 2 2 5 7 8
17	0 0 6 6 7
18	3 4 6 8
19	0 1 5 5
20	5
21	5

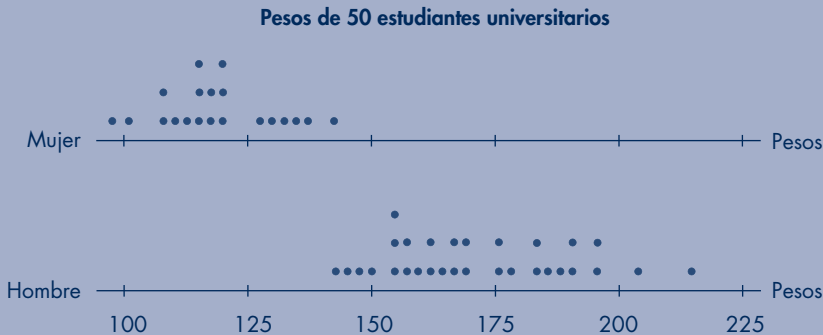
FIGURA 2.8  
Presentaciones de tallos y hojas “espalda con espalda”

Pesos de 50 estudiantes universitarios (lb)		
Mujeres		Hombres
8	09	
1 8 8	10	
0 2 5 5 6 8 8	11	
0 0 0 8 9	12	
2 5 7	13	
2	14	3 5 8
	15	0 4 4 5 7 8
	16	1 2 2 5 7 8
	17	0 0 6 6 7
	18	3 4 6 8
	19	0 1 5 5
	20	5
	21	5



La figura 2.9, una gráfica de puntos “lado a lado” (misma escala) de los mismos 50 datos de peso, muestra la misma distinción entre los dos subconjuntos. Con base en la información que se muestra en las figuras 2.8, 2.9 y en lo que se sabe acerca del peso de las personas, parece razonable concluir que las estudiantes universitarias pesan menos que los estudiantes universitarios. En el capítulo 3 se estudian las situaciones que involucran más de un conjunto de datos.

FIGURA 2.9  
Gráficas de puntos con escala común



INSTRUCCIONES DE TECNOLOGÍA:  
GRÁFICA DE PUNTOS MÚLTIPLES

MINITAB

Escribe los datos en C1 y las correspondientes categorías numéricas en C2; después continúa con:

Elige: **Graph > Dotplot . . .**  
Selecciona: **One Y, With Groups > OK**  
Escribe: **Graficar variables: C1**  
**Variables categóricas para agrupamiento: C2 > OK**

Si las diversas categorías están en columnas separadas, selecciona Multiple Y's Simple e ingresa todas las columnas bajo Graficar variables.

Excel

No están disponibles gráficas de puntos múltiples, pero puedes hacer el paso inicial de clasificar los datos. Usa los comandos como se muestran con la gráfica de puntos de la página 37; después termina la construcción de la gráfica de puntos a mano.

TI-83/84

Escribe los datos para la primera gráfica de puntos en L1 y los datos para la segunda gráfica de puntos en L3; después continúa con:

Elige: **STAT > EDIT > 2:SortA(**  
Escribe: **L1 > ENTER**  
En L2, escribe números de conteo para cada categoría.

Ej.	L1	L2
	15	1
	16	1
	16	2
	17	1

Elige: STAR > EDIT > 2:SortA(  
 Escribe: L3 > ENTER  
 En L4, escribe números de conteo (un conjunto\* superior) para cada categoría; \*por ejemplo: usa 10, 10, 11, 10, 10, 11, 12, ... (recorre las dos gráficas de puntos)

Elige: 2nd > FORMAT > AxesOff  
 (Opcional: debe regresar a AxesOn)

Elige: 2nd > STAT PLOT > 1:PLOT1

Elige: 2nd > STAT PLOT > 2:PLOT2

Elige: Window  
 Escribe: cuando mucho el valor más bajo para ambos, al menos el valor más alto para ambos, 0 o incremento, -2, al menos número de conteo más alto, 1, 1

Elige: Graph > Trace > > > >  
 (proporciona valores de datos)



## EJERCICIOS SECCIÓN 2.1

- 2.1** a. ¿Usualmente cuánto tiempo empleas en tu aseo por día?
- b. ¿Cómo crees que te comparas con los 50 000 estudiantes universitarios en “Estudiantes, aquí los observan” de la página 32?
- c. ¿Cómo crees que te comparas con todos los estudiantes universitarios? ¿Cuáles son las similitudes? ¿Cuáles son las diferencias?

**2.2** [EX02-002] A estudiantes en un curso de estadística en línea se les preguntó en cuántas diferentes actividades en internet se involucran durante una semana típica. Los siguientes datos muestran el número de actividades:

6	7	3	6	9	10	8	9	9	6	4	9	4	9
4	2	3	5	13	12	4	6	4	9	5	6	9	
11	5	6	5	3	7	9	6	5	12	2	6	9	

- a. Si se te pide presentar dichos datos, ¿cómo los organizarías y los resumirías?
- b. ¿En cuántas diferentes actividades en internet te involucraste la semana pasada?
- c. ¿Cómo crees que te comparas con los 40 usuarios de internet en la muestra anterior?

**2.3** Como gráfica estadística, la gráfica circular tiene limitaciones. Examina la gráfica circular de la figura 2.1 y la gráfica de barras en la figura 2.2.

- a. ¿Qué información muestran ambas?
- b. ¿Qué información se muestra en la gráfica circular que no se puede mostrar en la gráfica de barras?

- c. “En términos generales, la gráfica de barras es una mejor opción para usar que la gráfica circular.” Justifica esta afirmación.

**2.4** Los resultados de una encuesta Self.com acerca de “¿Cuál es tu principal preocupación de belleza en clima frío?”, se reportaron en el número de diciembre de 2008 de la revista *Self*: piel seca: 57%; labios agrietados: 25%; cabello sin brillo: 10%; pies ásperos: 8%.

- a. Construye una gráfica de pastel que muestre las principales preocupaciones de belleza de clima frío.
- b. Construye una gráfica de barras que muestre las principales preocupaciones de belleza de clima frío.
- c. En tu opinión, ¿la gráfica de pastel del inciso a o la gráfica de barras del inciso b resulta una mejor representación de la información? Explica.

**2.5** La American Payroll Association obtuvo una gran respuesta a esta pregunta acerca del código de vestido de la compañía: “El actual código de vestido en mi compañía es...”.

- Resultados finales:
- a. Demasiado relajado: 27%
  - b. Demasiado formal: 15%
  - c. Adecuado: 58%

La mayoría de las personas mencionó la importancia de la “comodidad” en sus explicaciones. La gran mayoría de los requeridos estuvo muy feliz con el código o política de vestido de su compañía.

- a. Construye una gráfica circular que muestre esta información. Etiquétala por completo.
- b. Construye una gráfica de barras que muestre esta misma información. Etiquétala por completo.

- c. Compara las dos gráficas anteriores y describe lo que ves en cada una ahora que las gráficas están completamente dibujadas y etiquetadas. ¿Obtienes la misma impresión acerca de los sentimientos de estas personas a partir de ambas gráficas? ¿Alguna enfatiza algo que la otra no?

**2.6** En la instantánea del *USA Today* del 5 de febrero de 2009, se reportó cuánto más los jóvenes estadounidenses, entre 17 y 28 años de edad, quieren pagar por un vehículo amigable con el ambiente: mucho más, 11%; un poco más, 36%; ligeramente más, 33%; no pagarían más, 20%.

- Menciona la variable de interés.
- Identifica el tipo de variable.
- Construye una gráfica de pastel que muestre cómo se sienten los jóvenes estadounidenses acerca de pagar por un vehículo amigable con el ambiente.
- Construye una gráfica de barras que muestre cómo se sienten los jóvenes estadounidenses acerca de pagar por un vehículo amigable con el ambiente.
- En tu opinión, ¿cuál gráfica es la mejor representación de la información? ¿Por qué? Explica.

**2.7** A continuación se muestra el número de puntos anotados por los equipos ganadores el 28 de octubre de 2008, la noche de apertura de la temporada 2008-2009 de la NBA:

Equipo	Boston	Chicago	LA Lakers
Puntos anotados	90	108	96

Fuente: <http://www.nba.com/>

- Dibuja una gráfica de barras de estas puntuaciones con una escala vertical que varíe de 80 a 110.
- Dibuja una gráfica de barras de las puntuaciones con una escala vertical que varíe de 50 a 110.
- ¿En cuál gráfica de barras parece que las puntuaciones de la NBA varían más? ¿Por qué?
- ¿Cómo podrías crear una representación precisa del tamaño relativo y la variación entre dichas puntuaciones?

**2.8 [EX02-008]** La American Community Survey recopila datos de estimaciones de población, demografía y unidades de alojamiento. Después la Oficina de Censos usa los datos para producir y disseminar estimaciones oficiales de unidades de alojamiento por estados y condados. A continuación se presentan las estimaciones de unidades de alojamiento 2005-2007 para la ciudad de Webster en el estado de Nueva York.

Unidades de alojamiento Webster, NY	
Unidades de alojamiento ocupadas por el propietario	12 627
Unidades de alojamiento ocupadas por arrendatario	3 803
Unidades de alojamiento vacantes	539
Total	16 969

Fuente: U.S. Census Bureau

- Construye una gráfica de pastel de este desglose.
- Construye una gráfica de barras de este desglose.
- Compara las dos gráficas que construiste en los incisos a y b; ¿cuál parece ser la más informativa? Explica por qué.

**2.9** Limpiar detrás de los muebles y lavar las ventanas encabezan la lista de labores domésticas de limpieza general, de acuerdo con la última Encuesta Nacional de Limpieza General de la Soap and Detergent Association (SDA). La International Communications Research (ICR) completó el estudio independiente de investigación del consumidor en enero/febrero de 2008. La pregunta inicial de la encuesta se planteó a 1 013 adultos estadounidenses (507 hombres y 506 mujeres).

La pregunta decía: ¿Regularmente se involucra en limpieza general?

Resultados: Sí, 77%  
No, 23%

Más mujeres (86%) que hombres (68%) hacen limpieza general.

- Construye y etiqueta completamente una gráfica de barras que muestre los resultados de todos los adultos encuestados.
- Construye y etiqueta completamente una gráfica de barras que muestre los resultados comparativos de mujeres y hombres por separado.
- Discute las gráficas de los incisos a y b y asegúrate de comentar acerca de con cuánta precisión, o no, las gráficas muestran la información.

Fuente: <http://www.cleaning101.com/>

**2.10 [EX02-010]** En ocasiones, las compañías de tarjetas de crédito brindan a sus consumidores un resumen al final del año. El resumen ofrece un reporte accesible y fácil de leer que resume las transacciones en varias categorías. Usa la tabla que aparece en la parte superior de la página 44:

- Explica el significado de las entradas de tabla de \$64.02 y \$(100.55).
- Explica el significado de los totales \$1 159.19 y \$298.42.
- Usa una gráfica de pastel para mostrar los totales de categoría a fin de año usando tanto cantidades en dólares como porcentajes. Asegúrate de etiquetar por completo.
- Usa una gráfica de barras para mostrar los totales mensuales. Asegúrate de etiquetar por completo.

**2.11** Un inspector de camisetas en una fábrica de ropa clasifica los últimos 500 defectos como: 67, falta botón; 153, mala costura; 258, tamaño inadecuado; 22, fallo de tela. Construye un diagrama de Pareto para esta información.



### Tabla para el ejercicio 2.10

Mes	Viaje	Restaurante	Mercancía	Auto	Servicios	Utilitarios	Totales
Enero	\$ —	\$ —	\$ 87.38	\$ —	\$ 13.80	\$ —	\$ 101.18
Febrero	\$ —	\$ 39.86	\$ 9.99	\$ 176.90	\$ (100.55)	\$ —	\$ 126.20
Marzo	\$ —	\$ 24.45	\$ —	\$ —	\$ 60.51	\$ —	\$ 84.96
Abril	\$ 25.00	\$ 135.78	\$ —	\$ —	\$ 260.00	\$ —	\$ 420.78
Mayo	\$ —	\$ —	\$ —	\$ —	\$ 175.27	\$ —	\$ 175.27
Junio	\$ 25.00	\$ 19.12	\$ 254.30	\$ —	\$ —	\$ —	\$ 298.42
Julio	\$ 25.00	\$ 46.94	\$ 281.12	\$ 64.02	\$ 30.00	\$ —	\$ 447.08
Agosto	\$ 25.00	\$ —	\$ 45.54	\$ —	\$ 21.48	\$ 35.40	\$ 127.42
Septiembre	\$ —	\$ 22.18	\$ —	\$ —	\$ 55.85	\$ —	\$ 78.03
Octubre	\$ 25.00	\$ 38.01	\$ —	\$ —	\$ 61.55	\$ —	\$ 124.56
Noviembre	\$ —	\$ —	\$ 86.51	\$ —	\$ 15.00	\$ —	\$ 101.51
Diciembre	\$ —	\$ —	\$ 394.35	\$ —	\$ 22.55	\$ —	\$ 416.90
Totales	\$ 125.00	\$ 326.34	\$ 1 159.19	\$ 240.92	\$ 615.46	\$ 35.40	\$ 2 502.31

**2.12 [EX02-012]** Las definiciones de correo electrónico spam, o correo electrónico basura, por lo general incluyen la idea de que el correo electrónico no es solicitado y se envía en masa. A principio de los años 1990, la cantidad de correo electrónico spam creció de manera constante hasta la actualidad, con un volumen total de más de 100 000 millones de correos electrónicos diarios en abril de 2008. La cantidad recibida comenzó a disminuir debido al uso de mejor software de filtrado. Por increíble que parezca, menos de 200 *spammers* enviaron alrededor de 80% de todo el spam.

El siguiente cuadro menciona los porcentajes de correo electrónico spam retransmitidos por cada país en 2007.

País	Porcentaje
Brasil	4.1
China	8.4
UE	17.9
Francia	3.3
Alemania	4.2
India	2.5
Italia	2.8
Polonia	4.8
Rusia	3.1
Corea del Sur	6.5
Turquía	2.9
Reino Unido	2.8
EUA	19.6

**Fuente:** <http://en.wikipedia.org/>

- Construye una gráfica de barras de esta información con los porcentajes en orden decreciente.
- Explica por qué no se puede construir un diagrama de Pareto de esta información.

**2.13** Un estudio completado por la International Communications Research para la Soap and Detergent Association (SDA) menciona el artículo que los estadounidenses dicen estarían más deseosos de ceder con la finalidad de poder contratar a alguien para hacer su limpieza general. La respuesta más popular fue \$100 (29%), seguido por cenar fuera durante un mes (26%), boletos para conciertos (19%), un viaje de fin de semana (9%) y otros (17%).

**Fuente:** <http://www.cleaning101.com/>

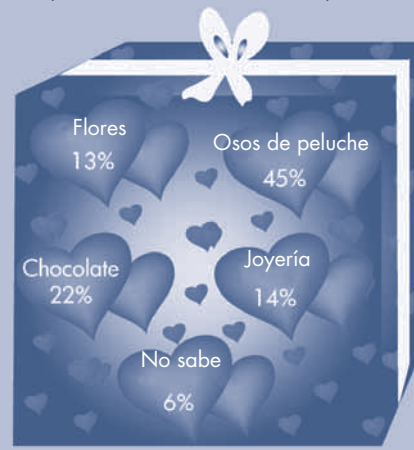
- Construye un diagrama de Pareto que muestre esta información.

- Debido al tamaño de la categoría “otros”, el diagrama de Pareto puede no ser la mejor gráfica a usar. Explica por qué y describe qué información adicional se necesita para hacer al diagrama de Pareto más apropiado.

### 2.14 ¿Qué NO dar el Día de san Valentín?

#### Presentes no deseados

Cuando se trata de regalos del Día de san Valentín, los adultos estadounidenses dicen que prefieren NO recibir osos de peluche.



**Fuente:** Datos tomados de Anne R. Carey y Juan Thomassie, *USA Today*

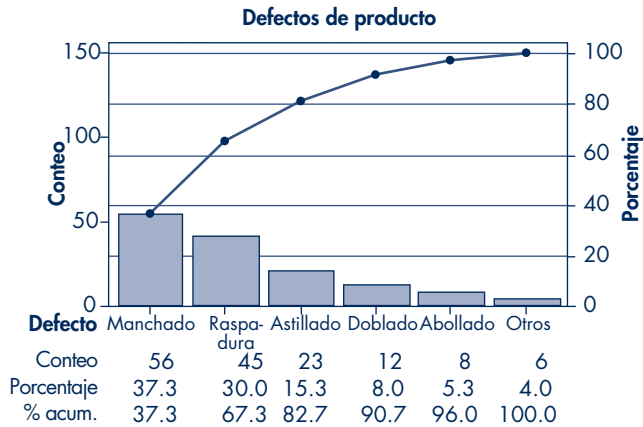
- Dibuja una gráfica de barras que muestre los porcentajes de “Presentes no deseados”.
- Dibuja un diagrama de Pareto que muestre los “Presentes no deseados”.
- Si quieres estar 80% seguro de no dar a tu ser amado algo que no quiere, ¿qué evitarías comprar? ¿Cómo muestra esto el diagrama de Pareto?
- Si se encuesta a 300 adultos, ¿qué frecuencias esperarías que ocurran para cada artículo no deseado mencionado en la gráfica?

**2.15** El reporte de defectos de la inspección final para la línea de ensamblado A12 se reporta en un diagrama de Pareto.

- ¿Cuál es el conteo de defecto total en el reporte?

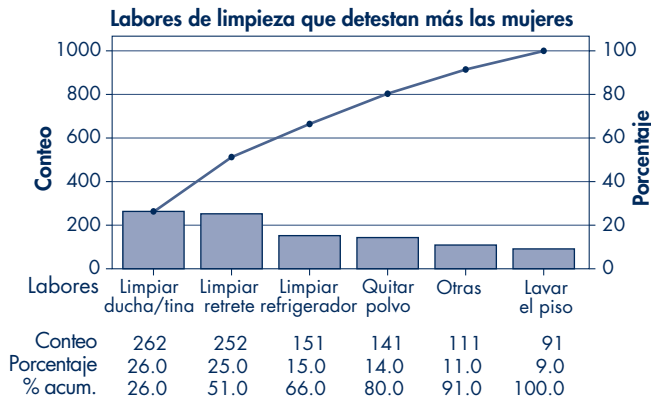


- b. Verifica el 30.0% mencionado para “Raspadura”.



- c. Explica cómo se obtuvo y qué significa el valor de 90.7% de “% acumulado para doblado”.
- d. La administración dio a la línea de producción la meta de reducir sus defectos en 50%. ¿A cuáles dos defectos sugerirías dar atención especial para trabajar hacia esta meta? Explica.

**2.16** Algunas labores de limpieza son más detestadas que otras. De acuerdo con la instantánea del *USA Today* del 17 de julio de 2009 acerca de una encuesta de mujeres del Consumer Reports National Research Center, las labores de limpieza que desagradan más a las mujeres se presentan en el siguiente diagrama de Pareto.



- a. ¿A cuántas mujeres en total se encuestó?
- b. Verifica el 15% mencionado para “Limpiar refrigerador”.
- c. Explica cómo se obtuvo y qué significa el valor de 80% para “% acumulado para quitar polvo”.
- d. ¿Cuáles tres labores harían felices a no más de 75% de las mujeres encuestadas, si dichas labores se eliminaran?

**2.17** [EX02-017] La American Time Use Survey que se presentó al comienzo del capítulo destacó el uso del tiempo de un día de la semana promedio para estudiantes de universidad de tiempo completo.

Categoría	Horas
Dormir	8.3
Ocio y deportes	3.9
Actividades educativas	3.2
Trabajo y actividades relacionadas	3.0
Comer y beber	1.0
Viajar	1.5
Aseo	0.8
Otro	2.3
Total	24.0

- a. Construye un diagrama de Pareto que muestre el uso de tiempo promedio para estudiantes universitarios de tiempo completo.
- b. ¿Qué actividades parecen constituir 75% del día de un estudiante universitario?

**2.18** [EX02-018] La Office of Aviation Enforcement and Proceedings, U.S. Department of Transportation, publicó esta tabla que menciona el número de quejas del consumidor contra las principales aerolíneas estadounidenses por categoría de queja.

Categoría de queja	Número	Categoría de queja	Número
Publicidad	68	Problemas de vuelo	2 031
Equipaje	1 421	Sobreventa	454
Servicio al cliente	1 715	Devoluciones	1 106
Discapacidad	477	Reservaciones/boletaje/abordaje	1 159
Tarifas	523	Otro	322

**Fuente:** Office of Aviation Enforcement and Proceedings, U.S. Department of Transportation, Air Travel Consumer Report, <http://www.infoplease.com/>

- a. Construye un diagrama de Pareto que muestre esta información.
- b. ¿En cuáles quejas recomendarías a las aerolíneas poner más atención para corregirlas si quieren tener el mejor efecto sobre el número global de quejas? Explica cómo el diagrama de Pareto del inciso a demuestra la validez de tu respuesta.

**2.19** [EX02-019] El número de puntos anotados durante cada juego por un equipo de baloncesto de bachillerato la última temporada fueron los siguientes: 56, 54, 61, 71, 46, 61, 55, 68, 60, 66, 54, 61, 52, 36, 64, 51. Construye una gráfica de puntos de dichos datos.

**2.20** [EX02-020] En un artículo del *USA Today* del 8 de julio de 2009, titulado “Parejas que dicen ‘no’ a bodas costosas”, los recortes pueden no extenderse al número de asistentes. En una encuesta de bodas recientes, el número de madrinas fue el siguiente:

7	6	5	2	3	7	6	13	6	3	2	7	8	9
---	---	---	---	---	---	---	----	---	---	---	---	---	---

- a. Construye un diagrama de puntos de dichos datos.
- b. ¿Cuáles son los números más comunes de madrinas? ¿Cómo muestra esto el diagrama de puntos?

**2.21 [EX02-021]** A continuación se muestran las alturas (en pulgadas) de los jugadores de baloncesto que fueron las primeras selecciones de los equipos profesionales de la National Basketball Association en 2009.

82	86	76	77	75	72	75	81	78	74
77	77	81	81	82	80	76	72	74	74
73	82	80	84	74	81	80	77	74	78

Fuente: <http://www.mynbadraft.com/>

- Construye una gráfica de puntos de las alturas de dichos jugadores.
- Usa la gráfica de puntos para descubrir a los jugadores más bajo y más alto.
- ¿Cuál es la altura más común y cuántos jugadores comparten dicha altura?
- ¿Qué característica de la gráfica de puntos ilustra la altura más común?

**2.22 [EX02-022]** La tabla menciona la mediana de los precios de venta de casas para los 20 suburbios de Rochester, Nueva York, según cita el *Democrat & Chronicle* del 18 de julio de 2009.

**Mediana de precios de casas en miles de dólares**

160	125	122	89	100	110	94	125	108	235
133	121	190	175	218	130	180	113	156	114

Fuente: Greater Rochester Association of Realtors

- Construye una gráfica de puntos de dichos datos.
- Describe la distribución que muestra la gráfica de puntos encontrada en el inciso a.

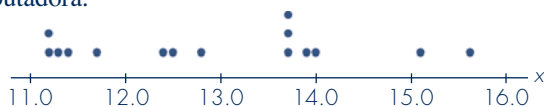
**2.23 [EX02-023]** Delco Products, una división de General Motors, produce conmutadores diseñados para tener una longitud total de 18.810 mm. (Un conmutador es un dispositivo que se usa en el sistema eléctrico de un automóvil.) La siguiente muestra de 35 longitudes de conmutador se tomó mientras se monitoreaba el proceso de fabricación:

18.802	18.810	18.780	18.757	18.824	18.827	18.825
18.809	18.794	18.787	18.844	18.824	18.829	18.817
18.785	18.747	18.802	18.826	18.810	18.802	18.780
18.830	18.874	18.836	18.758	18.813	18.844	18.861
18.824	18.835	18.794	18.853	18.823	18.863	18.808

Fuente: Con permiso de Delco Products Division, GMC

Usa una computadora para construir una gráfica de puntos de estos valores de datos.

**2.24** Para construir la siguiente gráfica de puntos se usó una computadora.



- ¿Cuántos valores de datos se muestran?
- Menciona los valores de los cinco datos más pequeños.
- ¿Cuál es el valor del objeto de datos más grande?

- ¿Qué valor ocurrió más número de veces? ¿Cuántas veces ocurrió?

**2.25 [EX02-025]** Construye una gráfica de tallo y hojas del número de puntos anotados durante cada juego de baloncesto la última temporada:

56	54	61	71	46	61	55	68
60	66	54	61	52	36	64	51

**2.26 [EX02-026]** En la tabla que se muestra a continuación se presentan las temperaturas máxima y mínima para cada una de 15 ciudades de México, de un día de octubre de 2011,

Ciudad	Temperatura mínima (°C)	Temperatura máxima (°C)
Acapulco	25	28
Aguascalientes	11	21
Campeche	23	28
Cd. de México	11	19
Cd. Juárez	13	30
Cd. Madero	24	31
Chihuahua	11	29
Guadalajara	12	24
Hermosillo	18	30
Ixtapa	23	29
Monterrey	18	38
Puebla	9	21
Querétaro	10	20
Tijuana	14	29
Zacatecas	8	21

- Construye el diagrama de tallos y hojas para la temperatura máxima y para la temperatura mínima.
- Con base en los diagramas anteriores describe la distribución de temperaturas máximas y de temperaturas mínimas.

**2.27 [EX02-027]** Las cantidades que se muestran a continuación son las tarifas que cobra Quik Delivery para los 40 paquetes pequeños que entregó el pasado jueves en la tarde:

4.03	3.56	3.10	6.04	5.62	3.16	2.93	3.82	4.30	3.86
4.57	3.59	4.57	6.16	2.88	5.03	5.46	3.87	6.81	4.91
3.62	3.62	3.80	3.70	4.15	2.07	3.77	5.77	7.86	4.63
4.81	2.86	5.02	5.24	4.02	5.44	4.65	3.89	4.00	2.99

- Construye un diagrama de tallo y hojas.
- Con base en el diagrama de tallo y hojas, describe la distribución de los datos.

**2.28 [EX02-028]** Una de las muchas cosas que reportó al público la U.S. Census Bureau es el aumento en población para varias áreas geográficas dentro del país. En la siguiente tabla se presenta el porcentaje de incremento en población para los 24 condados de más rápido crecimiento en Estados Unidos, del 1 de julio de 2006 al 1 de julio de 2007.

Condado	Estado	Porcentaje
St. Bernard Parish	Luisiana	42.9
Orleans Parish	Luisiana	13.8

\*\*\*Para el resto de los datos, ingresa en [cengagebrain.com](http://cengagebrain.com)

Fuente: <http://www.census.gov/>

- Construye un diagrama de tallo y hojas.
- Con base en el diagrama de tallo y hojas, describe la distribución de los datos.

**2.29** Dado el siguiente diagrama de tallo y hojas:

Steam-and-Leaf of C1 N = 16

Leaf Unit = 0.010

1	59	7
4	60	148
(5)	61	02669
7	62	0247
3	63	58
1	64	3

- ¿Cuál es el significado de “Leaf Unit (Unidad de hoja) = 0.010”?
- ¿Cuántos datos se muestran en este diagrama de tallo y hojas?
- Menciona los primeros cuatro valores de datos.
- ¿Qué es la columna de números a la izquierda de la figura?

**2.30** Un término que se usa con frecuencia en investigación en energía solar es *grados día de calefacción*. Este concepto

se relaciona con la diferencia entre una temperatura interior de 65 °F y la temperatura exterior promedio de un día dado. Una temperatura exterior promedio de 5 °F ofrece 60 grados día de calefacción. En el siguiente diagrama de tallo y hojas construido usando MINITAB, se muestran los días-grado de calefacción anuales normales para varias ubicaciones de Nebraska.

Steam-and-Leaf of C1 N = 25

Leaf Unit = 10

2	60	78
7	61	03699
9	62	69
11	63	26
(3)	64	233
11	65	48
9	66	8
8	67	249
5	68	18
3	69	145

- ¿Cuál es el significado de “Leaf Unit = 10”?
- Menciona los primeros cuatro valores de datos.
- Menciona todos los valores de datos que ocurrieron más de una vez.

## 2.2 Distribuciones de frecuencia e histogramas

Las listas de grandes conjuntos de datos no presentan una gran imagen. En ocasiones se quiere condensar los datos en una forma más manejable. Esto puede lograrse con la ayuda de una *distribución de frecuencias*.

**Distribución de frecuencias** Listado, con frecuencia expresado en forma de tabla, que relaciona los valores de una variable con su frecuencia.

**TABLA 2.4** Distribución de frecuencia no agrupada

$x$	$f$
0	1
1	3
2	8
3	5
4	3

Para demostrar el concepto de una distribución de frecuencia, utilicemos este conjunto de datos:

3	2	2	3	2	4	4	1	2	2
4	3	2	0	2	2	1	3	3	1

Si  $x$  representa la variable, entonces puedes usar una distribución de frecuencias para representar este conjunto de datos al hacer una lista de los valores  $x$  con sus frecuencias. Por ejemplo, el valor 1 ocurre en la muestra tres veces; por tanto, la **frecuencia** para  $x = 1$  es 3. En la tabla 2.4 se muestra el conjunto de datos completo en la distribución de frecuencias.

La frecuencia,  $f$ , es el número de veces que el valor  $x$  ocurre en la muestra. La tabla 2.4 es una **distribución de frecuencias no agrupadas**, “no agrupadas” porque cada valor de  $x$  en la distribución es independiente. Cuando un conjunto grande de datos tiene muchos valores  $x$  diferentes en lugar de algunos valores repetidos, como en el ejemplo anterior, puedes agrupar los valores en un conjunto de clases y construir una **distribución de frecuencias agrupadas**. El diagrama de tallo y hojas de la figura 2.5B (p. 38) muestra, en forma de imagen, una distribución de frecuencias agrupada. Cada tallo representa una clase. El número de hojas en cada tallo es el mismo que la frecuencia para dicha misma **clase** (en ocasiones llamada *caja*). Los datos que se presentan en la figura 2.5B se mencionan como una distribución de frecuencias agrupadas en la tabla 2.5.

TABLA 2.5 Distribución de frecuencias agrupadas

		Clase	Frecuencia
50 o más a menos de 60	—————>	$50 \leq x < 60$	1
60 o más a menos de 70	—————>	$60 \leq x < 70$	3
70 o más a menos de 80	—————>	$70 \leq x < 80$	8
80 o más a menos de 90	—————>	$80 \leq x < 90$	5
90 o más a menos de 100	—————>	$90 \leq x < 100$	2
			19

Puedes usar el proceso de tallo y hojas para construir una distribución de frecuencias; sin embargo, la representación en tallos no es compatible con todos los **anchos de clase**. Por ejemplo, los anchos de clase de 3, 4 y 7 son difíciles de usar. Por tanto, en ocasiones es ventajoso tener un procedimiento separado para construir una distribución de frecuencias agrupadas.

EJEMPLO 2.6



AGRUPAMIENTO DE DATOS PARA FORMAR UNA DISTRIBUCIÓN DE FRECUENCIAS

Para ilustrar este procedimiento de agrupamiento (o clasificación), usa una muestra de 50 calificaciones del examen final de la clase de estadística elemental del semestre pasado. La tabla 2.6 presenta las 50 calificaciones.

Procedimiento para construir una distribución de frecuencias agrupadas

1. Identifica la calificación alta ( $H = 98$ ) y la calificación baja ( $L = 39$ ) y encuentra el rango:
- $$\text{rango} = H - L = 98 - 39 = 59$$
2. Selecciona un número de clase ( $m = 7$ ) y un ancho de clase ( $c = 10$ ) de modo que el producto ( $mc = 70$ ) sea un poco mayor que el rango (rango = 59).

TABLA 2.6  
Calificaciones de examen de estadística [TA02-06]

60	47	82	95	88	72	67	66	68	98	90	77	86
58	64	95	74	72	88	74	77	39	90	63	68	97
70	64	70	70	58	78	89	44	55	85	82	83	
72	77	72	86	50	94	92	80	91	75	76	78	

3. Elige un punto de partida. Este punto de partida debe ser un poco menor que la calificación más baja,  $L$ . Supón que comienzas en 35; al contar desde las decenas (el ancho de clase), obtienes 35, 45, 55, 65, . . . , 95, 105. A ellos se les llama **límites de clase**. Las clases para los datos en la tabla 2.6 son:

35 o más a menos de 45	—————>	$35 \leq x < 45$
45 o más a menos de 55	—————>	$45 \leq x < 55$
55 o más a menos de 65	—————>	$55 \leq x < 65$
65 o más a menos de 75	—————>	$65 \leq x < 75$
	⋮	$75 \leq x < 85$
		$85 \leq x < 95$
95 o más a e incluido 105	—————>	$95 \leq x \leq 105$

Notas:

1. De un vistazo puedes verificar el patrón de número para determinar si la aritmética usada para formar las clases fue correcta (35, 45, 55, . . . , 105.)



- 2. Para el intervalo  $3.5 \leq x < 45$ , 35 es el límite de clase inferior y 45 es el límite de clase superior. Las observaciones que caen en el límite de clase inferior permanecen en dicho intervalo; las observaciones que caen en el límite de clase superior pasan al siguiente intervalo superior, excepto por la última clase.
- 3. El ancho de clase es la diferencia entre los límites de clase superior e inferior.
- 4. Cuando se clasifican datos, son posibles muchas combinaciones de anchos de clase, números de clases y puntos de partida. No hay una opción mejor. Intenta algunas combinaciones diferentes y usa el buen juicio para decidir la que usarás.

En consecuencia, se usan los siguientes **lineamientos básicos** para construir una distribución de frecuencia agrupada:

- 1. Cada clase debe ser del mismo ancho.
- 2. Las clases (en ocasiones llamadas *cajas*) deben establecerse de modo que no se traslapen y de modo que cada valor de dato pertenezca exactamente a una clase.
- 3. Para los ejercicios ofrecidos en este texto, de 5 a 12 clases es lo más deseable, porque todas las muestras contienen menos de 125 valores de datos. (La raíz cuadrada de  $n$  es un lineamiento razonable para el número de clases con muestras con menos de 125 valores de datos.)
- 4. Usa un sistema que saque ventaja de algún patrón para garantizar precisión.
- 5. Cuando sea conveniente, con frecuencia es ventajoso un ancho de clase par.

Una vez establecidas las clases, es necesario ordenar los datos en dichas clases. El método utilizado para ordenar dependerá del formato actual de los datos: si los datos están clasificados, las frecuencias pueden contarse; si los datos no están clasificados, **cuenta** los datos para encontrar los números de frecuencia. Cuando clasifiques datos, es útil usar un cuadro estándar (véase la tabla 2.7).

**TABLA 2.7**  
Cuadro estándar para distribución de frecuencias

Número de clase	Cuentas de clase	Límites	Frecuencia
1		$35 \leq x < 45$	2
2		$45 \leq x < 55$	2
3		$55 \leq x < 65$	7
4		$65 \leq x < 75$	13
5		$75 \leq x < 85$	11
6		$85 \leq x < 95$	11
7		$95 \leq x \leq 105$	4
			50

**Notas:**

- 1. Si los datos están clasificados (en forma de lista, gráfica de puntos o tallo y hojas), ya no es necesario clasificar; sólo cuenta los datos que pertenecen a cada clase.
- 2. Si los datos no están clasificados, ten cuidado con tu clasificación y conteo.
- 3. La frecuencia,  $f$ , para cada clase es el número de piezas de datos que pertenecen a dicha clase.
- 4. La suma de las frecuencias debe ser igual al número de piezas de datos,  $n$  ( $n = \sum f$ ). Esta suma sirve como una buena comprobación.